



Universidade de Brasília – UnB  
Instituto de Geociências - IG

DESENVOLVIMENTO DE MODELO ESTATÍSTICO  
ESPACIAL PARA ESTIMATIVA DA VARIAÇÃO DO NÍVEL  
FREÁTICO DE POÇOS PIEZOMÉTRICOS NA BACIA  
EXPERIMENTAL DO ALTO RIO JARDIM

LEANDRO DE ALMEIDA SALLES

MONOGRAFIA

Orientador: Dr. Jorge Enoch Furquim Werneck Lima

Co-orientador: Prof. Dr. Henrique Llacer Roig

Brasília – DF  
2012



Universidade de Brasília – UnB  
Instituto de Geociências - IG

DESENVOLVIMENTO DE MODELO ESTATÍSTICO  
ESPACIAL PARA ESTIMATIVA DA VARIAÇÃO DO NÍVEL  
FREÁTICO DE POÇOS PIEZOMÉTRICOS NA BACIA  
EXPERIMENTAL DO ALTO RIO JARDIM

Leandro de Almeida Salles

Monografia apresentada junto ao curso de Pós-graduação  
em Geoprocessamento Ambiental, do Instituto de  
Geociências da Universidade de Brasília para obtenção  
do título de especialista em Geoprocessamento.

Orientador: Dr. Jorge Enoch Furquim Werneck Lima

Co-orientador: Prof. Dr. Henrique Llacer Roig

Brasília – DF  
2012



Universidade de Brasília – UnB  
Instituto de Geociências - IG

DESENVOLVIMENTO DE MODELO ESTATÍSTICO  
ESPACIAL PARA ESTIMATIVA DA VARIAÇÃO DO NÍVEL  
FREÁTICO DE POÇOS PIEZOMÉTRICOS NA BACIA  
EXPERIMENTAL DO ALTO RIO JARDIM

LEANDRO DE ALMEIDA SALLES

Banca examinadora:

Dr. Jorge Enoch Furquim Werneck Lima (Embrapa)

Prof. Dr. Paulo Roberto Meneses (IG – UnB)

Prof. Dra. Tati de Almeida (IG - UnB)

Brasília – DF  
2012

“Omolu, Ogum, Oxum, Oxumaré,  
    Todo o pessoal,  
    Manda descer pra ver!  
Iansã, Iemanjá, chama Xangô  
    Oxossi também  
    Manda descer pra ver!  
Mercador, Cavaleiro de Bagdá  
    Oh, Filhos de Obá  
Manda descer pra ver:  
    Filhos de Gandhi!”  
    **Gilberto Gil**

***DEDICATÓRIA***

*À minha esposa Adriana.*

*Aos meus pais Paulo e Heloísa.*

*Aos meus avós Zilmar e Bento.*

*Aos meus avós Tinda e Flávio.*

*Amo vocês!*

## AGRADECIMENTOS

Ao meu orientador Dr. Jorge Werneck pela confiança no meu trabalho, pelos ensinamentos ao longo deste período, pela disponibilização de dados e pelas horas dedicadas à elaboração desta monografia.

Ao meu co-orientador Prof. Dr. Henrique Llacer Roig pelos ensinamentos ao longo deste período de convivência.

Ao Analista Estatístico da Embrapa Juaci Malaquias pelo auxílio técnico na elaboração deste trabalho.

Ao Prof. Dr. Paulo Roberto Meneses e à Prof. Dra. Tati de Almeida por aceitarem participar da banca examinadora.

Gostaria de agradecer minha querida esposa Adriana pelo companheirismo e carinho ao longo de mais uma caminhada e etapa da minha vida.

Aos meus pais, Paulo e Heloísa que sempre me apoiaram, me amaram e sempre foram a fonte de inspiração em todos os aspectos da minha vida.

Aos meus irmãos e irmãs, aos meus avôs e avós, às minhas sobrinhas, às minhas tias e tios, primos, primas e demais membros da minha família pelo amor e amizade que nos une e por cada dia de nossas vidas que estivemos juntos.

Ao Instituto Brasília Ambiental por investir na minha formação profissional e aos amigos de lá que sempre me apoiaram.

Ao povo brasileiro que através da estrutura existente na Universidade de Brasília, no Instituto de Geociências me proporcionou a experiência de cursar a Pós-Graduação *lato sensu* em Geoprocessamento Ambiental.

Aos professores de todas as disciplinas do curso de Pós-Graduação *lato sensu* em Geoprocessamento Ambiental pelos ensinamentos. Ao Professor Paulo Roberto Meneses, Serjão e Eluzai pelo auxílio nas questões administrativas ligadas ao curso de Pós-Graduação *lato sensu* em Geoprocessamento Ambiental.

Aos amigos Felipe Meirelles, Vicente Arcela e Paulo Henrique pelas boas discussões e ensinamentos sobre o assunto estudado.

Aos amigos da UnB pela convivência acadêmica e profissional que sempre trouxeram boas reflexões.

Aos demais amigos pela companhia ao longo desta caminhada.

## SUMÁRIO

1. INTRODUÇÃO .....	1
2. OBJETIVOS .....	4
2.1 Objetivo Geral .....	4
2.2 - Objetivo Específico .....	4
3. MATERIAIS E MÉTODOS.....	4
3.1 Área de estudo .....	4
3.2 Variável dependente .....	6
3.3 Variável independente .....	7
3.4 Regressão linear múltipla.....	10
3.5 Regressão Com Pesos Espaciais.....	12
3.6 Programas computacionais .....	15
4. RESULTADOS E DISCUSSÃO.....	15
4.1 Análise de correlação.....	15
4.2 Regressão linear global .....	18
4.3 Regressão com pesos espaciais .....	25
4.4 Validação dos modelos de regressão global e com pesos espaciais .....	32
5. CONCLUSÕES .....	38
6. CONSIDERAÇÕES FINAIS .....	39
Referências Bibliográficas .....	39

## ÍNDICE DE FIGURAS

Figura 1: Localização da bacia do Alto Rio Jardim e dos 21 poços piezométricos utilizados para construção do modelo estatístico e 14 poços piezométricos utilizados para validação do modelo.....	5
Figura 2: Relação entre os resíduos e os valores ajustados pela regressão do modelo com maior parcimônia (seis variáveis explicativas). ....	23
Figura 3: Gráfico QQnorm dos resíduos do modelo com maior parcimônia (seis variáveis explicativas).....	23
Figura 4: Distribuição espacial dos coeficientes estimados pela regressão com pesos espaciais. ....	27
Figura 5: Amplitude máxima de variação do lençol freático na bacia do Alto Rio Jardim observada e estimada pela regressão com pesos espaciais, e respectivo resíduo do modelo (a), (b) e (c). ....	29
Figura 6: $R^2$ local e $R^2$ local ajustado do modelo gerado pela regressão com pesos espaciais para estimar a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim, (a) e (b). ....	31
Figura 7: Correlação entre a amplitude máxima da variação do lençol freático observada nos poços piezométricos selecionados para validação localizados na bacia do Alto Rio Jardim e a estimada pelo modelo de regressão linear global (a) e com pesos espaciais (b).....	33
Figura 8: Resíduos obtidos na validação dos modelos desenvolvidos para simular a amplitude máxima da variação do lençol freático pela regressão global (a) e com pesos espaciais e (b). ....	34
Figura 9: Correlação entre a amplitude máxima da variação do lençol freático observada nos poços piezométricos selecionados para validação, retirando-se os poços nº10 e nº13 localizados na bacia do Alto Rio Jardim e a estimada pelo modelo de regressão linear global (a) e com pesos espaciais (b). ....	36
Figura 10: Correlação entre a amplitude máxima da variação do lençol freático observada nos poços piezométricos selecionados para validação, retirando-se os poços nº10, nº13 e nº20 localizados na bacia do Alto Rio Jardim e a estimada pelo modelo de regressão com pesos espaciais. ....	37



## ÍNDICE DE TABELAS

Tabela 1: Código das variáveis para construção da matriz de correlação.....	9
Tabela 2: Matriz de correlação de Pearson da amplitude máxima da variação do nível freático dos poços piezométricos da bacia do Alto rio Jardim no período entre 17/02/2004 e 10/01/2012, e das variáveis físicas do solo e geomorfológicas da bacia do Alto rio Jardim. ....	16
Tabela 3: Resultado dos coeficientes estimados na regressão para as variáveis selecionadas para compor o full model. ....	19
Tabela 4: Resultado dos coeficientes estimados na regressão para as variáveis selecionadas para compor o modelo selecionado pela regressão <i>stepwise</i> no sentido backward. ....	20
Tabela 5: Resultado dos coeficientes estimados na regressão para as variáveis selecionadas para compor o modelo com maior parcimônia (seis variáveis explicativas). ....	22
Tabela 6: Sumário da estimativa dos coeficientes obtidos na regressão com pesos espaciais. ....	25
Tabela 7: ANOVA comparando os resultados do modelo de regressão global e com pesos espaciais. ....	31

## RESUMO

O objetivo deste trabalho foi avaliar, por meio de análise de regressão linear multivariada global e de técnica de geoprocessamento ambiental que realiza regressão com pesos espaciais, a correlação entre a amplitude máxima da variação do nível freático dos poços piezométricos da bacia do Alto rio Jardim no período entre 17/02/2004 e 10/01/2012, e os fatores ambientais relacionados às propriedades físicas do solo e a variáveis geomorfológicas presentes na bacia. Do total de 35 variáveis explicativas disponíveis, foram selecionados seis variáveis para compor o modelo final de regressão global e com pesos espaciais: teor de argila, teor de silte, densidade das partículas do solo, densidade média do solo, e porosidade do solo a 60 cm de profundidade, além da densidade média do solo na camada superficial. A regressão com pesos espaciais, por sua vez, foi conduzida utilizando a função de pesos espaciais bi-quadrado, com *kernel* espacial adaptativo, buscando otimizar o *Akaike Information Criterion* “corrigido”, AICc. A regressão global foi considerada significativa pelo teste F e os resultados de diagnóstico tiveram  $R^2$ ,  $R^2$  ajustado, e AICc, respectivamente, iguais a 0,84, 0,73, e 71,52. A análise dos resíduos indicou comportamento homocedástico pelo teste Koenker; que a variação dos resíduos se assemelha à distribuição normal pelos testes de Jarque-Bera e Shapiro-Wilks; que não há resíduos considerados como *outliers* pelo teste de Bonferonni; e, por fim, que os resíduos não apresentam autocorrelação espacial segundo os resultados obtidos pelo Índice de Moran. A regressão com pesos espaciais foi considerada significativa pelo teste F e os resultados de diagnóstico  $R^2$ ,  $R^2$  ajustado e AICc foram, respectivamente, iguais a: 0,98; 0,90; e -1.220,71. Os resíduos da regressão com pesos espaciais não apresentam autocorrelação espacial segundo os resultados do Índice de Moran. Os resultados demonstram que a regressão com pesos espaciais conseguiu captar correlações locais das variáveis explicativas com a variável resposta que não foram obtidas pela regressão linear global, indicando benefícios da técnica de geoprocessamento na análise ambiental. Porém, não houve diferença significativa entre os métodos utilizados, de modo que, ambos os métodos foram capazes de estabelecer correlações entre a variável resposta e a variável explicativa.

Palavras Chave: Regressão pesos espaciais, Poços Piezométricos.

## ABSTRACT

The purpose of this study was to analyze the correlations between environmental factors, such as physical soil proprieties and geomorphology, with the maximum water level variation of piezometric wells, from 17/02/2004 to 10/01/2012, in the upper Jardim experimental river basin using Ordinary Least Squares Regression, OLS, and Geographic Weighted Regression, GWR. Six out of 35 explanatory variables available were selected to compose the final OLS model and the GWR model: amount of silt and clay from the 60cm soil layer; soil particles density from the 60cm soil layer; bulk density at the surface and at the 60cm layer; soil porosity at the 60cm layer. The GWR was done using bi-square weighting function with adaptive kernels optimizing the AICc. The F-test considered the OLS significant. The diagnostic statistic for the OLS was equal to 0.84; 0.73; 59.52 and 71.52 for the  $R^2$ ,  $R^2_{adjusted}$ , AIC and AICc, respectively. The OLS residual analysis showed that at a 5% significance level, the selected model's residuals had an homocedastic behavior (KoenkerBP's test), similar to the normal distribution (Jarque-Bera's test and Shapiro-Wilks' test), without any outliers (Bonferonni's test), and with no spatial autocorrelation (Moran's I). The F-test considered the GWR significant. The diagnostic statistic for the GWR was equal to 0.98; 0.90; 52 and -1220.71 for the  $R^2$ ,  $R^2_{adjusted}$ , AIC and AICc, respectively. There was no spatial autocorrelation (Moran's I) on the GWR's residuals at a 5% significance level. The GWR model was capable to find local correlation that the OLS model could not find, showing benefits of the geoprocessing technique on environmental analysis. However, the results indicate that there was no significant difference between the GWR and OLS results, with both methods being capable of establishing correlations between the response variable and the explanatory variables.

Keywords: Geographic Weighted Regression, Piezometric wells

# 1. INTRODUÇÃO

A gestão dos recursos hídricos é um assunto multidisciplinar que possui diferentes metodologias que podem ser empregadas. Os modelos de gerenciamento dos recursos hídricos podem ser classificados como modelos de: otimização, cujo objetivo é buscar as melhores soluções de um sistema específico, em nível de projeto; planejamento, os quais simulam condições globais de um sistema maior e buscam além de soluções hidráulicas, hidrológicas e econômicas, quantificações ambientais e socioeconômicas; e comportamento que, por sua vez, são utilizados para o prognóstico da resposta de sistemas sujeito a diferentes entradas, ou com diferentes características (Tucci, 1998).

A maneira em que as entradas do modelo se relacionam com as suas saídas pode classificar os modelos de gestão dos recursos hídricos em determinísticos e não determinísticos.

Nos modelos determinísticos, o impacto das perturbações atribuídas à aleatoriedade dos parâmetros no sistema é ignorado e, além disso, o sistema físico como ele ocorre, em um dado grupo de eventos, pode ser governado por um resultado identificável, enquanto, nos modelos não determinísticos o sistema físico é definido com resultado incerto e aleatório, podendo ser expresso em termos probabilísticos ou estocásticos (Chow, 1988).

Em sistemas determinísticos, a mesma entrada, *input*, irá sempre produzir a mesma saída, podendo o *input* ocorrer de forma determinística ou aleatória, enquanto em sistemas probabilísticos sempre haverá pelo menos um elemento cuja relação do *input* com o *output* será probabilística (Dooge, 1973).

Os processos hidrológicos são geralmente contínuos no tempo e no espaço, onde as variáveis que quantificam os fenômenos assumem valores ao longo de qualquer ponto, temporal ou espacialmente. Não obstante, devido aos métodos disponíveis para fazer modelagem, é comum representar os processos hidrológicos por variáveis aleatórias temporais discretizadas, sendo que o termo discretizado significa que são tratados os valores das variáveis obtidos em instantes ou períodos sucessivos do tempo, ou em pontos geográficos definidos no espaço (Lanna, 2001).

O uso das técnicas de análise multivariada pode fornecer informações interessantes acerca da distribuição espacial e das tendências temporais de dados hidrológicos (Bierman et al., 2011), sendo ferramentas importantes para o manejo adequado dos recursos hídricos.

As técnicas de análise multivariada são empregadas em diferentes áreas do conhecimento para se conhecer a estrutura e as relações de dados multivariados. Características físicas de rochas sedimentares, como distribuição granulométrica, porosidade, permeabilidade; variáveis fluviais como descarga, material em suspensão, profundidade; e bandas espectrais em imagens de satélites são exemplos comuns da existência de dados multivariados em Geociências (Landim, 2010).

Utilizando um conjunto de dados do sensor MODIS associado a um banco de dados de dados de clorofila-a mensurados no Golfo de Spencer no sul da Austrália, e com base em exemplos encontrados na literatura, Bierman *et al.* (2011), explica como diferentes métodos de análise multivariada podem ser aplicados no estudo de dados hidrológicos, agrupando-os em três classes com as seguintes técnicas:

- Classificação: análise de agrupamentos (*Cluster analysis*) que classifica objetos em grupos, hierárquicos ou não, com mesmas características; análise discriminante que busca determinar variáveis que podem segregar objetos em grupos por meio de metodologia *stepwise*.
- Sumarização: análise de componentes principais que identifica as fontes de variação de dados multivariados; análise fatorial que identifica os fatores responsáveis pela variação de dados espaciais; e *self-organizing maps* cuja aplicação está relacionada à busca de cenários, ou de padrões espaciais de dados multivariados com distribuição espacial.
- Análise de relações espaciais: Semivariograma, utilizado para compreender relações espaciais entre dados medidos e como método de interpolação; regressão com pesos espaciais cujo objetivo está relacionado em compreender mudanças na relação de variáveis no espaço.

Algumas dessas técnicas não levam em consideração a localização dos pontos de amostragem, assim como as relações espaciais existentes entre os dados medidos como,

por exemplo, a análise discriminante, e a análise de componentes principais, enquanto, outras buscam entender se a organização espacial é significativa, como a krigagem (Landim, 2010) e a regressão com pesos espaciais (Fotheringham *et al.*, 2002).

A regressão com pesos espaciais pode ser utilizada como um método mais acurado do que a regressão linear clássica: por incorporar na estrutura do modelo a variação das relações entre diferentes as variáveis causadas pela distribuição espacial; e como uma maneira de entender, de um modo diferenciado, a variação espacial (Bierman *et al.*, 2011). Todavia, a incorporação do elemento espacial na modelagem estatística de regressões lineares pode não trazer ganhos significativos ao processo (Fotheringham *et al.*, 2002) cabendo uma análise criteriosa sobre os resultados.

Na regressão com pesos espaciais, os parâmetros são estimados para cada ponto de coleta de dados, levando em consideração o peso dado para as observações de acordo com a sua proximidade do ponto em questão, de modo que quanto mais próximas estiverem as observações vizinhas deste ponto, maior será o peso (Fotheringham *et al.*, 2002). Com isso, é possível estudar localmente as diferentes correlações entre as variáveis em questão.

A regressão com pesos espaciais é uma técnica que busca correlacionar espacialmente variáveis e tem sido amplamente utilizada ultimamente. Assim como as técnicas de análise multivariada mencionadas, esse método pode ser empregado em diferentes áreas do conhecimento como em estudos epidemiológicos (Leyk *et al.*, 2012; Gilbert & Chakraborty, 2011), em estudos com dados derivados de sensoriamento remoto (Bayramov *et al.*, 2012; Erdogan, 2010; Gao *et al.*, 2012), em estudos com dados sócios econômicos (Breetzke, 2012; Deller, 2011; Grubestic *et al.*, 2012; Himmelberger *et al.*, 2009), estudos florestais (Kupfer & Farris, 2007), e em estudos de transporte (Blainey, 2010; Ibeas *et al.*, 2011). Também há exemplos de estudos relacionados à qualidade da água que fizeram uso dessa técnica (Harris & Brunsdon, 2010; Mills *et al.*, 2009; Wu *et al.*, 2009).

Em bacias hidrográficas do Cerrado brasileiro, onde o período de chuva dura apenas cerca de seis meses, o conhecimento sobre o comportamento do fluxo da água no solo é fundamental, pois é esta a fonte que mantém a disponibilidade hídrica nos rios da região durante o período seco do ano. O monitoramento da variação espaço-temporal do nível

freático possibilita identificar quanta água é armazenada nessa região do solo e onde ela se encontra, fornecendo ainda informações importantes sobre zonas de recarga da bacia e, se for o caso, regiões favoráveis para instalação de cisternas para o aproveitamento direto desse recurso hídrico. Por se tratar de informação complexa e de difícil levantamento em campo, a busca por dados obtidos próximo à superfície do terreno que expliquem essas características do lençol freático pode ser de grande utilidade. Diante do exposto, este estudo buscou compreender relações entre as características físicas e geomorfológicas da bacia experimental do Alto Rio Jardim (DF) com a variação do nível do lençol freático da bacia, utilizando técnicas de estatística multivariada e geoprocessamento.

## **2. OBJETIVOS**

### **2.1 Objetivo Geral**

Avaliar, por meio de técnica espacial e não espacial de análise multivariada, a correlação entre o comportamento do lençol freático da bacia hidrográfica do Alto Rio Jardim no período entre 17/02/2004 e 10/01/2012 e fatores ambientais existentes na respectiva bacia.

### **2.2 - Objetivo Específico**

- Avaliar a correlação entre a amplitude máxima de variação do nível do lençol freático na bacia do Alto Rio Jardim e os fatores ambientais existentes na respectiva bacia.
- Ajustar modelo de regressão global para predição da amplitude máxima de variação do nível do lençol freático na bacia do Alto Rio Jardim.
- Ajustar modelo de regressão com pesos espaciais para predição da amplitude máxima de variação do nível do lençol freático na bacia do Alto Rio Jardim.

## **3. MATERIAIS E MÉTODOS**

### **3.1 Área de estudo**

A bacia hidrográfica do Alto Rio Jardim localiza-se no Distrito Federal, Figura 1, entre as latitudes 15,71° e 15,86°S, e as longitudes 47,55° e 47,64°W. A área de drenagem da bacia hidrográfica do Alto Rio Jardim é aproximadamente 104,86 km², e engloba duas sub-bacias: a do Córrego Estanislau e a do rio Jardim, com 49,71 e 55,15km², respectivamente (Lima, 2010). Em relação às grandes regiões hidrográficas brasileiras, a bacia do rio Jardim está inserida na região do rio São Francisco, sendo afluente do rio Preto que por sua vez, é afluente do rio Paracatu o qual é um importante contribuinte da margem esquerda do rio São Francisco.

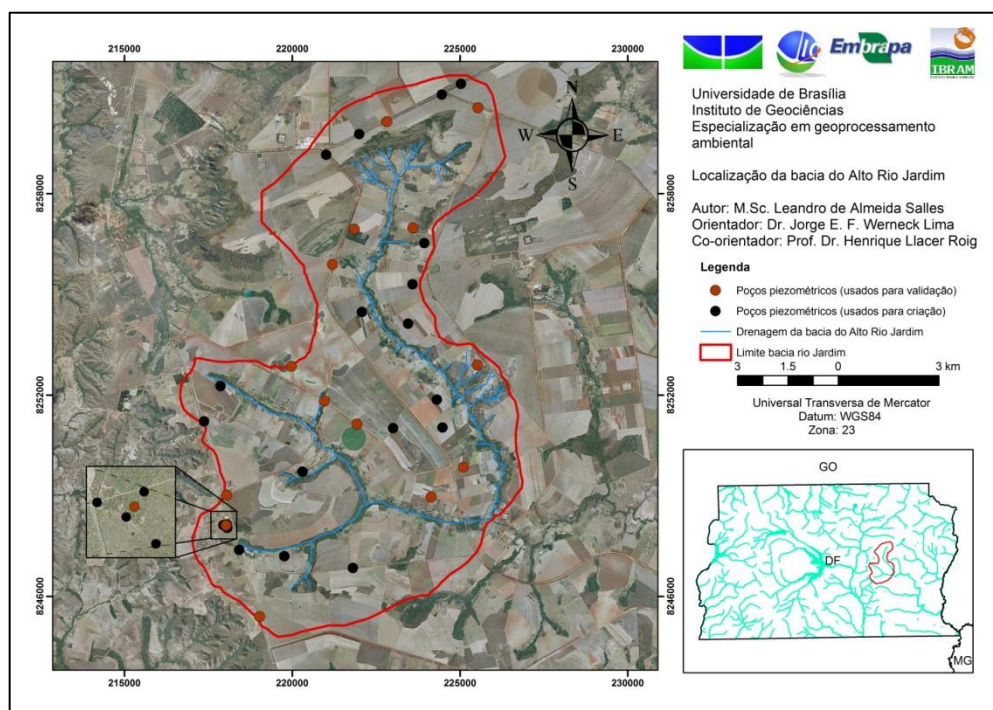


Figura 1: Localização da bacia do Alto Rio Jardim e dos 21 poços piezométricos utilizados para construção do modelo estatístico e 14 poços piezométricos utilizados para validação do modelo.

O clima na bacia hidrográfica do Alto Rio Jardim possui sazonalidade na distribuição das chuvas, apresentando duas estações bem definidas: inverno seco entre os meses de abril e outubro; e verão chuvoso nos demais meses do ano com precipitação média anual variando entre 1.000mm a 1.700mm, e com mais de 80% das chuvas ocorrendo neste período. De acordo com a classificação proposta por Köppen, o clima



predominante na região é do tipo Tropical de Savana (Codeplan, 1984 *apud* Lima 2010).

A partir da análise da Imagem Landsat ETM+, de 18 de Julho de 2003, da região na qual está inserida a bacia hidrográfica do Alto Rio Jardim, Lima *et al.* (2007) classificou o uso do solo na bacia da seguinte maneira: 76,4% ocupado para fins agropecuários, divididos por pastagens, e agricultura irrigada ou de sequeiro; 22,4% ocupado por vegetação nativa, sendo 14,9% de Cerrado e 7,5% de Mata de Galeria ou Ciliar; aproximadamente 1,0% ocupado por solo exposto e edificações; e o restante, 0,2%, ocupado por corpos d'água. A bacia hidrográfica do Alto Rio Jardim localiza-se em região agrícola do Distrito Federal e as principais atividades agropecuárias desenvolvidas são o cultivo de grãos (soja, feijão, milho, sorgo), algodão, cítricos, café, mandioca, hortaliças, e criação de aves e de gado (Lima, 2010).

Os tipos de solos que ocorrem na bacia hidrográfica do Alto Rio Jardim, segundo levantamento semi-detalhado na escala 1:50.000 feito por Reatto *et al.* (1998) são classificados em: Latossolo com 76,38% da área total; Cambissolo com 16,68% da área total; Plintossolo com 2,54% da área total; Gleissolo representando 2,41% da área total; e Neossolo Quartzarênico com 2,09% do total. Além desses tipos de solos, também ocorrem afloramento de rocha com aproximadamente 0,24% da área total.

### **3.2 Variável dependente**

A variável dependente (variável resposta) utilizada no presente estudo foi a amplitude máxima da variação do nível freático dos poços piezométricos da bacia do Alto rio Jardim no período entre 17/02/2004 e 10/01/2012. As informações do nível freático dos poços piezométricos foram coletadas mensalmente neste período.

O comportamento hidrológico dos poços piezométricos utilizados na análise não é homogêneo, com a amplitude máxima do nível freático variando entre 1,85 e 8,49m. Lima (2010) avaliou graficamente a relação entre a precipitação registrada em uma estação pluviométrica e a variação da profundidade dos poços piezométricos, todos localizados dentro da bacia do Alto Rio Jardim, demonstrando que no ano hidrológico

2007/2008 a variação do nível da água nos poços piezométricos monitorados apresentou diferentes profundidades, amplitudes e velocidades de resposta em relação à chuva.

Foram utilizados dados de 21 poços piezométricos para construção do modelo linear de regressão global e com pesos espaciais e dados de 14 poços para validar os resultados de ambos os modelos desenvolvidos. Na Figura 1 é apresentada a distribuição espacial dos 35 poços piezométricos usados na análise. Todos os poços foram numerados aleatoriamente utilizando ferramenta disponível no *software Microsoft Excel*. Posteriormente, utilizando a mesma ferramenta, foram sorteados quais poços piezométricos seriam utilizados para o desenvolvimento do modelo e para a sua validação.

Vale ressaltar que houve quatro exceções nos poços utilizados para a validação dos modelos, os quais não participaram do sorteio. Esses poços não tinham disponíveis na base de dados todos os dados das variáveis explicativas que seriam utilizados para construção do modelo de regressão. Porém, após o desenvolvimento do modelo observou-se que eles possuíam os dados das variáveis explicativas que foram selecionados na regressão global. Assim, optou-se por utilizar esses quatro poços para a validação do modelo.

### 3.3 Variável independente

As variáveis independentes utilizadas para o desenvolvimento do modelo de regressão linear, ou seja, as variáveis que buscam explicar o comportamento da amplitude máxima do nível freático dos poços piezométricos da bacia do Alto Rio Jardim foram coletadas por Lima (2010) no mesmo local em que os poços piezométricos foram instalados. O banco de dados coletado por Lima (2010) é composto, principalmente, por variáveis que representam as propriedades físicas dos solos nos diferentes locais.

Também foram obtidas variáveis geomorfológicas da bacia as quais foram derivadas das informações presente na carta SICAD por meio de técnicas de geoprocessamento, e da coleta de dados por DGPS feita por Lima (2010). Vale ressaltar que foi utilizado o sistema de projeção UTM (*Universal Transverse Mercator*), zona 23S, com *datum*

WGS1984 como o sistema padrão para elaboração do trabalho, devido aos resultados obtidos pelo DGPS na coleta das coordenadas e da altitude de cada poço piezométrico.

O banco de dados da Embrapa (Lima, 2010) possui as seguintes propriedades físicas do solo a serem testadas como variáveis explicativas para construção do modelo: Capacidade de água disponível no solo ( $\text{mm.mm}^{-1}$ ) nas profundidades 15 e 60 cm; teor de silte (%), argila (%), areia fina (%) e areia grossa (%) na superfície do solo e nas profundidades 15 e 60cm; teor de matéria orgânica (%) na superfície do solo e nas profundidades 15 e 60cm; Condutividade hidráulica saturada ( $\text{m.s}^{-1}$ ) a 15 e 60cm de profundidade; condutividade hidráulica do lençol freático ( $\text{m.s}^{-1}$ ); Densidade das partículas do solo ( $\text{g.m}^{-3}$ ) e densidade média do solo ( $\text{g.m}^{-3}$ ) na superfície do solo e nas profundidades 15 e 60cm; Porosidade do solo ( $\text{cm}^3.\text{cm}^{-3}$ ) a 15 e 60cm de profundidade; e a umidade do solo submetida a uma tensão de 10kPa ( $\text{g.g}^{-1}$ ) a 15 e 60cm de profundidade.

As variáveis geomorfológicas disponíveis para serem utilizadas como variáveis explicativas na construção do modelo eram: a altitude elipsoidal (m) de cada poço piezométrico; a distância mínima dos poços piezométricos em relação à rede de drenagem (m); a distância dos poços piezométricos em relação ao exutório da bacia do Alto Rio Jardim (m); a orientação dos poços piezométricos em relação ao norte geográfico; e a declividade (%) no local em que os poços piezométricos foram furados.

Desse modo, haviam 30 variáveis relacionadas às propriedades físicas do solo e cinco variáveis geomorfológicas, totalizando 35 variáveis disponíveis para a construção do modelo linear que explica a relação entre a amplitude máxima da variação do nível freático dos poços piezométricos da bacia do Alto rio Jardim no período entre 17/02/2004 e 10/01/2012.

Todavia, a existência de multicolinearidade entre as variáveis explicativas pode acarretar problemas inferenciais para o estimador de mínimos quadrados, pois as variâncias das estimativas de coeficientes e preditores podem crescer exageradamente mascarando a significância das variáveis utilizadas no modelo e limitando a aplicabilidade e a amplitude de previsão do modelo linear ajustado (Souza, 1998). Além disso, como o número de incógnitas, ou variáveis explicativas, era maior do que os 21 pontos amostrais selecionados para a construção do modelo linear, seria

matematicamente impossível encontrar uma solução possível para o sistema de equações lineares.

Assim, foi construída a matriz de correlação de Pearson para avaliar a relação entre todas as variáveis e assim, seleccionar aquelas variáveis que poderiam melhor expressar a relação entre a amplitude máxima da variação do nível freático dos poços piezométricos e os fatores ambientais. Tendo em vista o número de variáveis avaliadas, fez-se necessário estabelecer um código para cada variável para melhor visualizar a matriz de correlação conforme exposto na Tabela 1.

Tabela 1: Código das variáveis para construção da matriz de correlação.

Código	Nome	Código	Nome	Código	Nome
V.Dep	Amplitude máxima da variação do lençol freático	V12	Densidade das partículas do solo (15 cm)	V24	Teor de areia fina (15 cm)
V1	Altitude elipsoidal	V13	Densidade das partículas do solo (60 cm)	V25	Teor de areia fina (60 cm)
V2	Capacidade de Água disponível (15 cm)	V14	Teor de matéria orgânica (camada superficial do solo)	V26	Teor de areia grossa (camada superficial do solo)
V3	Capacidade de Água disponível (60 cm)	V15	Teor de matéria orgânica (15 cm)	V27	Teor de areia grossa (15 cm)
V4	Porosidade do solo (15cm)	V16	Teor de matéria orgânica (60 cm)	V28	Teor de areia grossa (60 cm)
V5	Condutividade hidráulica do solo (15 cm)	V17	Teor de argila (camada superficial do solo)	V29	Densidade média do solo (camada superficial do solo)
V6	Umidade do solo a 10kPa (15 cm)	V18	Teor de argila (15 cm)	V30	Densidade média do solo (15 cm)
V7	Porosidade do solo (60 cm)	V19	Teor de argila (60 cm)	V31	Densidade média do solo (60 cm)
V8	Condutividade hidráulica do solo (60 cm)	V20	Teor de silte (camada superficial do solo)	V32	Distância mínima do poço até a rede de drenagem
V9	Umidade do solo a 10kPa (60 cm)	V21	Teor de silte (15 cm)	V33	Orientação da vertente (°)
V10	Condutividade hidráulica do lençol freático	V22	Teor de silte (60 cm)	V34	Declividade (%)
V11	Densidade das partículas do solo (camada superficial do solo)	V23	Teor de areia fina (camada superficial do solo)	V35	Distância do poço até o exutório da bacia

### 3.4 Regressão linear múltipla

O modelo geral para análise de regressão linear múltipla pode ser expressa pela equação a seguir:

$$Y_i = \beta_0 + \sum_k \beta_k X_{ik} + \varepsilon_i \quad (1)$$

em que  $Y_i$  é o valor da variável dependente  $Y$  no ponto  $i$ ,  $\beta_0$  é o ponto em que a reta da regressão cruza o eixo das ordenadas,  $\beta_k$  é o coeficiente de declividade da  $k$ -ésima variável independente,  $X_{ik}$  é o valor da  $k$ -ésima variável independente no ponto  $i$ , e o  $\varepsilon_i$  representa o erro aleatório independente no ponto  $i$  com distribuição  $N(0, \sigma^2)$ .

A proporção em que a variância da variável resposta consegue ser expressa pela equação estimada pela regressão é representada pelo  $R^2$  pela seguinte equação:

$$R^2 = \frac{SQR}{SQT} \quad (2)$$

em que  $R^2$  é o coeficiente de determinação,  $SQR$  é a soma dos quadrados dos resíduos e  $SQT$  é a soma dos quadrados total. O  $R^2$  pode ser ajustado em função dos graus de liberdade do modelo pela equação:

$$R_{ajustado}^2 = 1 - \frac{n-1}{n-m} (1 - R^2) \quad (3)$$

em que  $n$  é o número de observações e  $m$  é o número de variáveis explicativas do modelo mais um.

O modelo foi selecionado utilizando a técnica *stepwise regression* na direção *backward* disponível no pacote R-Comander (Fox, 2005). Segundo Landim (2010), a regressão *stepwise* é uma das técnicas. Neste método, inicialmente é feita a análise de regressão com todas as  $k$  variáveis explicativas selecionadas (*full model*). Na etapa seguinte, o algoritmo testa o modelo retirando cada variável independente, uma por vez, e com base no *Akaike Information Criterion* (AIC) o algoritmo avalia os ganhos na retirada de cada variável. Caso seja mais vantajoso retirar uma variável, o algoritmo retira do modelo àquela variável explicativa cuja saída cause a menor redução no AIC, restando  $k-1$  variáveis. O algoritmo repete esta última etapa até o momento em que a retirada de uma variável não seja mais vantajosa do que a permanência das variáveis que ainda restam no modelo.

Para garantir que o modelo linear desenvolvido tenha significância estatística fez-se necessário que os resultados obtidos na regressão fossem submetidos a uma análise criteriosa para evitar a utilização inapropriada das técnicas estatísticas clássicas (Souza, 1998). Assim, a análise residual foi conduzida utilizando as seguintes técnicas:

- Teste F para avaliar a significância do modelo. A hipótese nula,  $H_0$ , deste teste representa a situação em que as variáveis explicativas do modelo não são efetivas.
- Teste Koenker (*Koenker's studentized Breusch-Pagan statistic*) para avaliar a estacionariedade do modelo. Neste teste a hipótese nula,  $H_0$ , representa a situação em que o modelo é estacionário. O teste avalia a relação entre as variáveis explicativas e a variável resposta tanto no espaço geográfico quanto no conjunto de dados, ou seja, se as variáveis explicativas se comportam da mesma maneira no espaço, e se a variação das respostas do modelo se mantém constante com a mudança na magnitude das variáveis explicativas (homocedasticidade), respectivamente.
- Teste de Bonferonni para análise de *outliers*, sendo que a hipótese nula,  $H_0$ , deste teste assume que não há *outliers* nos resíduos.
- Teste de Shapiro-Wilks e de Jarque-Bera para avaliar a normalidade dos resíduos. Em ambos os teste, a hipótese nula,  $H_0$ , assume que a distribuição de probabilidade da variável aleatória de interesse se assemelha à distribuição normal. Segundo Souza (1998), o teste de Shapiro-Wilks é um dos mais potentes e é efetivo para amostras relativamente pequenas ( $n \leq 50$ ).
- Índice de Moran para análise de autocorrelação espacial global dos resíduos. Nesse teste, a hipótese nula,  $H_0$ , indica que a variável de interesse se distribui aleatoriamente no espaço. Foi utilizado o inverso da distância como referência da relação espacial entre as feições. Nesse método, considera-se que todas as feições se influenciam, porém aquelas feições que estão mais próximas exercem maior influência do que as feições mais distantes, de modo que o peso é dado no inverso da distância. Como cada feição deve ter pelo menos um vizinho para que seja feito o cálculo do Índice de Moran, a distância limiar utilizada para busca dos vizinhos foi igual a 2.577,91m.

Também foi avaliado para cada variável explicativa o Fator Multiplicador da Variância, VIF. Este índice é importante por ser uma medida do impacto multiplicador da colinearidade na variância do coeficiente estimado ( $\hat{\beta}_k$ ), sendo que a prática estatística sugere que valores maiores do que 10 indicam a existência de multicolinearidade (Souza, 1998). Neste estudo, optou-se por aceitar  $VIF \leq 15$ , tendo em vista que também foi utilizada a análise da correlação linear de Pearson para verificar a correlação entre as variáveis explicativas utilizadas no modelo de regressão linear.

### 3.5 Regressão com pesos espaciais

O modelo geral para análise de regressão com pesos espaciais, por sua vez, pode ser expresso pela equação a seguir:

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)X_{ik} + \varepsilon_i \quad (4)$$

em que  $(u_i, v_i)$  é a coordenada do  $i$ -ésimo ponto no espaço e  $\beta_k(u_i, v_i)$  é a realização da função  $\beta_k(u, v)$  no ponto  $i$ , e o  $\varepsilon_i$  representa o erro aleatório independente no ponto  $i$  com distribuição  $N(0, \sigma^2)$ . Observa-se que a Equação (4) é um caso especial da (1) em que se assume que os parâmetros da regressão não variam no espaço (Fotheringham *et al.*, 2002).

Os parâmetros da regressão em cada ponto  $i$  observado podem ser calculados resolvendo a seguinte equação matricial:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (5)$$

em que  $\hat{\beta}$  representa a estimativa de  $\beta$ , e  $W(u_i, v_i)$  é uma matriz diagonal de dimensão  $n \times n$ , cujos elementos da diagonal,  $w_{in}$ , são os pesos espaciais para cada uma das  $n$  observações para o ponto  $i$  da regressão, e o restante dos elementos da matriz é igual a zero. Segundo Fotheringham *et al.* (2002), a Equação (5) assemelha-se ao estimador dos mínimos quadrados com a diferença de que no método dos mínimos quadrados o peso é constante, enquanto na regressão com pesos espaciais os pesos variam de acordo com a localização do ponto  $i$ .

Neste estudo foi utilizada a função de pesos espaciais bi-quadrado, em que os pesos são calculados a partir das equações a seguir:

$$w_{ij} = \left\{ 1 - \left( \frac{d_{ij}}{b} \right)^2 \right\}^2 \quad \text{para } d_{ij} < b \quad (6)$$

$$w_{ij} = 0 \quad \text{para } d_{ij} \geq b \quad (7)$$

em que  $i$  é o ponto em que o parâmetro da regressão está sendo estimado,  $j$  representa um ponto específico no espaço que os dados vizinhos ao ponto  $i$  foram observados,  $w_{ij}$  é o peso espacial,  $d$  é a distância, e  $b$  é a largura da banda (*bandwidth*) do *kernel* espacial utilizado para busca dos vizinhos do ponto  $i$ .

Optou-se por utilizar o *kernel* espacial que varia de acordo com a densidade dos pontos vizinhos do  $i$ -ésimo ponto, de modo que a largura da banda  $b$  não é um valor constante. Assim, o uso da função de pesos espaciais bi-quadrado com *kernel* espacial que varia no espaço requer que seja estabelecido o número de vizinhos do  $i$ -ésimo ponto para serem calibrados pela função de peso espacial, conforme as Equações (6) e (7), de modo que seja atribuído o peso zero para as observações que estiverem além desse limiar e o valor obtido pela função bi-quadrado para quem estiver dentro do limite (Fotheringham *et al.*, 2002). O critério de inclusão estabelecido neste estudo foi que o *kernel* espacial contemplasse entre 10 e 45% das  $n$  observações vizinhas do  $i$ -ésimo ponto.

Além do critério de inclusão dos vizinhos do  $i$ -ésimo ponto no *kernel* espacial para o uso da função de peso bi-quadrado, também foi necessário estabelecer o critério de calibração desta função. Neste estudo, optou-se pelo método de minimização do *Akaike Information Criterion* (AIC). Segundo Hurvich *et al.* (1998) *apud* Fotheringham *et al.* (2002) o AIC para a regressão com pesos espaciais é calculado pela equação a seguir:

$$AIC_c = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left\{ \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right\} \quad (8)$$

em que  $n$  é o tamanho da amostra,  $\hat{\sigma}$  é o desvio padrão do erro, e  $\text{tr}(S)$  é o traço da matriz *hat* a qual é função da largura da banda utilizada no *kernel* espacial. O subscrito  $c$  para no termo AICc denota que o resultado do *Akaike Information Criterion* foi “corrigido”. Vale ressaltar que os resultados do AIC e do AICc não podem ser comparados (Segundo Fotheringham *et al.*, 2002).



O *Akaike Information Criterion* (AIC) é calculado a partir da seguinte equação:

$$AIC = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n + \text{tr}(S) \quad (9)$$

em que  $n$  é o tamanho da amostra,  $\hat{\sigma}$  é o desvio padrão do erro, e  $\text{tr}(S)$  é o traço da matriz *hat* a qual é função da largura da banda utilizada no *kernel* espacial. O desvio padrão do erro é calculado a partir da seguinte equação:

$$\hat{\sigma} = \frac{SQR}{n} \quad (10)$$

em que SQR é a soma dos quadrados dos resíduos baseada na máxima verossimilhança. A matriz *hat*,  $S$ , por sua vez, é calculada pela seguinte equação matricial:

$$\hat{Y} = SY \quad (11)$$

em que  $\hat{Y}$  é o valor da variável dependente estimada pela regressão, e cada linha de  $S$ ,  $r_i$ , é calculado pela expressão:

$$r_i = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) \quad (12)$$

Os resíduos do modelo com pesos espaciais foram avaliados em termos absolutos de acordo com a equação a seguir:

$$Res = y_{obs} - y_{est} \quad (13)$$

em que  $Res$  é o resíduo,  $y_{obs}$  e  $y_{est}$  são, respectivamente, o valor observado e o valor estimado pelo modelo, de modo que resíduos negativos indicam que a resposta foi superestimada pelo modelo.

Os resultados dos coeficientes estimados para a constante do modelo e para as variáveis explicativas, do  $R^2$  local, da variável resposta estimada e dos resíduos foram interpolados pelo método IDW (*Inverse Distance Weighted*), com potência quadrática, para poder avaliar a variabilidade espacial dos resultados. Foram definidos 12 pontos no raio de busca, o qual foi permitido variar de acordo com a densidade de pontos, e célula de 5m que corresponde a uma escala 1:25.000. A superfície interpolada pelo método IDW foi recortada para a área da bacia do Alto Rio Jardim.

### 3.6 Programas computacionais

Para elaboração deste trabalho foram utilizados os seguintes sistemas computacionais:

- Software Arc-GIS 9.3; com as ferramentas, *Spatial Analyst*; e *Spatial Statistics* desenvolvidos pela Enviromental Systems Research Institute (ESRI), Redlands, CA, EUA.
- Aplicativo R, desenvolvido pela R Core Team (2012), e o pacote R-Comander desenvolvido por Fox (2005).
- Spatial Analysis in Macroecology (SAM), desenvolvido por Rangel *et al.* (2010).

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Análise de correlação

A Tabela 2 apresenta a matriz de correlação de Pearson das variáveis disponíveis para construção do modelo de regressão linear. A correlação linear entre a amplitude máxima da variação do nível freático dos poços piezométricos da bacia do Alto Rio Jardim (variável resposta) no período entre 17/02/2004 e 10/01/2012 e as variáveis explicativas variou entre 0,54 e -0,59.

A porosidade do solo a 60cm de profundidade (V7), a densidade das partículas do solo a 60cm de profundidade (V13), o teor de silte a 15cm de profundidade (V21), a densidade das partículas do solo na camada superficial do solo (V11) e o teor de silte a 60cm de profundidade (V22) apresentaram, respectivamente, as maiores correlações positivas com a variável resposta: 0,54; 0,50; 0,50; 0,43; 0,39; e 0,39. Já a condutividade hidráulica do solo a 60cm de profundidade (V8), o teor de silte na camada superficial (V20), a umidade do solo a 15cm de profundidade (V6) e a densidade média do solo na camada superficial do solo (V29) apresentaram correlações positivas com a variável resposta, respectivamente, iguais a 0,27, 0,23, 0,20 e 0,19.

Tabela 2: Matriz de correlação de Pearson da amplitude máxima da variação do nível freático dos poços piezométricos da bacia do Alto rio Jardim no período entre 17/02/2004 e 10/01/2012, e das variáveis físicas do solo e geomorfológicas da bacia do Alto rio Jardim.

	V.Dep.	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32	V33	V34	V35
V.Dep.	1,00																																			
V1	-0,59	1,00																																		
V2	-0,02	-0,10	1,00																																	
V3	-0,04	0,13	0,45	1,00																																
V4	-0,01	-0,04	-0,52	-0,23	1,00																															
V5	-0,46	0,39	-0,28	0,16	0,38	1,00																														
V6	0,20	-0,67	0,63	0,13	-0,08	-0,31	1,00																													
V7	0,54	-0,48	-0,31	-0,17	0,48	-0,13	0,18	1,00																												
V8	0,27	-0,51	-0,13	-0,72	0,06	-0,18	0,20	0,25	1,00																											
V9	0,02	-0,11	0,36	0,80	0,09	0,14	0,38	-0,11	-0,55	1,00																										
V10	-0,33	0,27	-0,18	0,00	0,44	0,32	-0,14	-0,24	-0,25	0,14	1,00																									
V11	0,39	-0,37	0,46	0,30	-0,45	-0,49	0,35	-0,01	0,07	0,25	-0,47	1,00																								
V12	0,13	0,03	0,24	0,38	-0,25	-0,24	-0,03	-0,06	-0,20	0,23	-0,22	0,51	1,00																							
V13	0,50	-0,28	0,25	0,36	-0,55	-0,42	0,24	-0,05	-0,20	0,22	-0,29	0,55	0,29	1,00																						
V14	-0,21	-0,10	-0,27	-0,27	0,09	0,35	-0,06	-0,12	0,26	-0,31	0,03	-0,28	-0,57	-0,07	1,00																					
V15	0,09	-0,43	0,10	-0,35	-0,06	-0,10	0,37	-0,12	0,49	-0,26	0,02	0,04	-0,17	0,16	0,51	1,00																				
V16	0,12	-0,36	0,29	-0,31	-0,23	-0,19	0,44	-0,10	0,55	-0,22	-0,23	0,12	-0,11	0,26	0,24	0,67	1,00																			
V17	-0,09	0,19	-0,37	-0,14	0,54	0,15	-0,14	0,40	-0,16	-0,03	0,06	-0,07	0,22	-0,30	-0,11	-0,28	-0,45	1,00																		
V18	-0,27	0,25	-0,25	0,01	0,57	0,15	-0,02	0,27	-0,24	0,23	0,22	-0,23	0,11	-0,39	-0,29	-0,45	-0,42	0,77	1,00																	
V19	-0,27	0,26	-0,43	-0,38	0,56	0,19	-0,23	0,16	0,09	-0,13	0,21	-0,24	-0,05	-0,51	-0,11	-0,29	-0,35	0,73	0,75	1,00																
V20	0,23	-0,56	0,01	-0,19	0,01	-0,07	0,35	0,08	0,52	0,04	0,05	-0,01	-0,55	0,03	0,28	0,33	0,35	-0,64	-0,35	-0,26	1,00															
V21	0,43	-0,69	-0,20	-0,39	0,23	-0,04	0,33	0,36	0,62	-0,13	-0,05	0,05	-0,57	0,02	0,51	0,49	0,33	-0,24	-0,37	-0,15	0,75	1,00														
V22	0,39	-0,70	-0,04	0,02	0,17	0,01	0,48	0,39	0,29	0,21	-0,01	-0,01	-0,43	0,14	0,37	0,36	0,23	-0,31	-0,24	-0,45	0,71	0,78	1,00													
V23	-0,08	0,37	0,46	0,39	-0,69	-0,15	-0,20	-0,57	-0,40	0,03	-0,15	0,20	0,39	0,38	-0,26	-0,04	0,13	-0,52	-0,56	-0,61	-0,31	-0,53	-0,40	1,00												
V24	-0,10	0,35	0,45	0,36	-0,75	-0,17	-0,22	-0,58	-0,36	-0,04	-0,17	0,24	0,39	0,38	-0,23	0,00	0,08	-0,52	-0,60	-0,60	-0,29	-0,50	-0,41	0,99	1,00											
V25	-0,03	0,31	0,49	0,38	-0,73	-0,22	-0,15	-0,51	-0,35	-0,01	-0,22	0,28	0,39	0,42	-0,24	-0,02	0,14	-0,53	-0,59	-0,67	-0,28	-0,48	-0,34	0,98	0,98	1,00										
V26	-0,30	0,08	-0,16	-0,10	0,18	0,27	-0,06	0,06	0,07	-0,18	0,14	-0,51	-0,25	-0,28	0,43	0,00	0,08	0,17	0,16	0,22	-0,15	-0,05	-0,08	-0,24	-0,27	-0,29	1,00									
V27	-0,11	0,05	-0,21	-0,13	0,03	0,33	-0,21	0,07	0,25	-0,31	0,04	-0,34	0,00	-0,17	0,31	0,04	0,11	0,05	-0,09	0,09	-0,16	-0,07	-0,18	-0,05	-0,06	-0,08	0,80	1,00								
V28	-0,06	0,07	0,03	0,09	-0,04	0,04	-0,05	0,07	0,08	-0,16	0,00	-0,12	0,15	0,05	0,23	0,13	0,21	0,10	-0,05	-0,05	-0,30	-0,17	-0,19	0,05	0,03	0,06	0,76	0,81	1,00							
V29	0,19	0,16	0,35	-0,18	-0,56	-0,46	-0,13	-0,30	0,07	-0,41	-0,19	0,37	0,14	0,18	-0,19	0,19	0,11	-0,24	-0,51	-0,28	-0,17	-0,10	-0,32	0,56	0,62	0,59	-0,41	-0,26	-0,16	1,00						
V30	0,08	0,03	0,44	0,18	-0,81	-0,43	-0,08	-0,34	-0,05	-0,18	-0,43	0,46	0,34	0,44	-0,10	0,18	0,26	-0,51	-0,67	-0,60	-0,08	-0,16	-0,18	0,74	0,80	0,78	-0,29	-0,16	0,03	0,70	1,00					
V31	-0,34	0,59	0,36	0,29	-0,54	0,05	-0,32	-0,84	-0,31	0,07	0,04	0,09	0,33	0,16	-0,11	-0,02	0,12	-0,38	-0,31	-0,26	-0,32	-0,60	-0,59	0,81	0,78	0,75	-0,06	0,09	0,16	0,42	0,51	1,00				
V32	0,04	0,33	-0,19	0,03	0,23	0,07	-0,27	0,03	-0,37	0,12	0,00	-0,15	-0,22	-0,08	0,02	-0,23	-0,17	0,21	0,12	-0,01	-0,21	0,05	0,07	0,02	-0,05	0,00	-0,21	-0,49	-0,33	0,11	0,02	-0,02	1,00			
V33	0,09	0,28	0,09	-0,09	-0,22	0,07	-0,12	-0,24	-0,17	-0,17	-0,05	-0,01	-0,18	0,27	0,16	0,08	0,16	-0,22	-0,41	-0,38	-0,14	-0,01	-0,08	0,46	0,41	0,50	-0,27	-0,09	-0,14	0,39	0,22	0,35	0,37	1,00		
V34	-0,09	-0,09	0,28	-0,05	-0,18	-0,08	0,33	0,22	0,14	-0,19	0,05	-0,07	0,03	0,15	-0,17	0,03	0,29	-0,03	0,10	0,08	0,03	-0,14	-0,10	-0,06	-0,03	-0,05	0,31	0,32	0,33	-0,10	0,02	-0,17	-0,52	-0,23	1,00	
V35	-0,57	0,92	-0,14	0,00	0,02	0,35	-0,65	-0,40	-0,33	-0,21	0,26	-0,38	0,01	-0,45	-0,14	-0,45	-0,42	0,25	0,30	0,42	-0,51	-0,66	-0,77	0,23	0,23	0,17	0,20	0,25	0,19	0,14	-0,08	0,52	0,09	0,14	-0,04	1,00

A altitude (V1), a distância dos poços piezométricos até o exutório (V35), a condutividade hidráulica a 15cm de profundidade (V5), a densidade média do solo a 60cm de profundidade (V31), a condutividade hidráulica do lençol freático (V10), e o teor de areia grossa na camada superficial do solo (V26) apresentaram, respectivamente, as maiores correlações negativas com a variável resposta: -0,59; -0,57; -0,46; -0,34; -0,33.

Além dessas variáveis explicativas, também vale citar o teor de argila a 60cm de profundidade (V19), o teor de argila a 15cm de profundidade (V18) e o teor de matéria orgânica na camada superficial do solo (V14) que tiveram correlações negativas com a variável resposta, respectivamente, iguais a: -0,27; -0,27; e -0,21.

Conforme exposto anteriormente, a correlação entre a variável resposta e as variáveis explicativas é importante, principalmente, na seleção das variáveis que serão utilizadas para a construção do modelo, não obstante, é necessário que seja feita uma análise criteriosa da relação entre as variáveis explicativas, de modo que não sejam escolhidos variáveis redundantes, multicolineares (Souza, 1998; Mitchell, 2005).

Os resultados demonstram que algumas das variáveis explicativas apresentam altíssima correlação entre elas, principalmente quando se trata da mesma variável nas diferentes camadas do solo.

O teor de argila na camada superficial (V17) teve correlação igual a 0,77 e 0,73 com o teor de argila a 15cm e a 60cm de profundidade (V18 e V19), respectivamente, e a correlação entre o teor de argila nestas últimas profundidades foi igual a 0,75. Em relação ao teor de silte, as correlações foram iguais a: 0,75 entre a camada superficial (V20) e a 15cm de profundidade (V21); 0,71 entre a camada superficial e a 60cm de profundidade (V22); e 0,78 entre as duas camadas mais profundas.

As correlações entre o teor de areia fina na camada superficial do solo (V23) e a 15cm e 60cm de profundidade (V24 e V25), foram, respectivamente, iguais a 0,99 e 0,98, e a correlação entre o teor de areia fina nas camadas mais profundas foi igual 0,98. Já as correlações entre o teor de areia grossa na camada superficial do solo (V26) e a 15cm e 60cm de profundidade (V27 e V28), e entre estas duas últimas profundidades foram, respectivamente, iguais a: 0,80; 0,76; e 0,81.

Quanto às variáveis geomorfológicas, observa-se que a maior correlação ocorreu entre a altitude (V1) e a distância dos poços ao exutório da bacia (V35), sendo igual a 0,92. A declividade (V34) apresentou correlação igual a -0,52 com a distância mínima do poço até a rede de drenagem (V32). Desse modo, com base nos resultados da Tabela 2 e na interpretação dos processos físicos e hidrológicos associados à variação do nível freático, às variáveis físicas do solo e às variáveis geomorfológicas, foram selecionadas 13 variáveis explicativas para compor o *full model*, do total de 35 variáveis.

#### 4.2 Regressão linear global

Os resultados da análise de regressão para o *full model* teve  $R^2$  igual 0,87,  $R^2$  ajustado para 7 graus de liberdade igual a 0,62, e AIC igual a 66,47. O teste F para esse modelo apresentou p-valor igual a 0,0511, indicando que a hipótese nula de que todas as variáveis explicativas são iguais pode ser rejeitada ao nível de significância de 10% e não ao nível de 5%.

A Tabela 3 apresenta os resultados dos coeficientes estimados pela análise de regressão do modelo com as 13 variáveis explicativas selecionadas, *full model*. Observa-se que apenas o coeficiente que cruza o eixo das ordenadas e a variável Porosidade do solo a 60cm de profundidade apresentaram resultado do p-valor menor do 0,05. Em relação ao índice de inflação da variância (VIF), observa-se que a densidade média do solo a 60cm de profundidade excedeu 15, indicando que esta variável apresenta multicolinearidade com as outras variáveis explicativas.

Os resultados da análise de regressão *stepwise* no sentido *backward*, cujo objetivo foi selecionar um modelo mais adequado do que o *full model* para explicar a amplitude máxima variação do lençol freático na bacia do Alto Rio Jardim no período entre 17/02/2004 e 10/01/2012, teve  $R^2$  igual 0,86,  $R^2$  ajustado para 11 graus de liberdade igual a 0,75, e AICc igual a 59,52. O teste F para esse modelo apresentou p-valor igual a 0,0014, indicando que a hipótese nula de que todas as variáveis explicativas são iguais pode ser rejeitada ao nível de significância de 1%.

Tabela 3: Resultado dos coeficientes estimados na regressão para as variáveis selecionadas para compor o full model.

Variáveis explicativas	Estimativa do coeficiente	Erro Padrão	t-valor	p-valor	VIF
Constante	-74,84	29,51	-2,54	0,04	-----
Teor de argila (60cm)	0,11	0,06	1,86	0,11	6,32
Distância do poço até o exutório da bacia	-0,000185	0,00	-0,92	0,39	6,58
Densidade das partículas do solo (camada superficial do solo)	-0,72	3,62	-0,20	0,85	2,68
Densidade das partículas do solo (60cm)	5,57	2,98	1,87	0,10	2,86
Densidade média do solo (camada superficial do solo)	6,32	3,36	1,88	0,10	2,56
Densidade média do solo (60cm)	24,09	11,40	2,11	0,07	15,41
Condutividade hidráulica do solo (15cm)	-0,02156	0,26	-0,08	0,94	2,36
Condutividade hidráulica do solo (60cm)	0,29	0,57	0,50	0,63	2,06
Condutividade hidráulica do lençol freático	0,060	0,41	0,15	0,89	1,88
Teor de matéria orgânica (camada superficial do solo)	-0,40	0,40	-1,00	0,35	2,13
Porosidade do solo (60cm)	63,14	22,27	2,84	0,03	7,47
Teor de silte (60cm)	0,19	0,09	2,14	0,07	9,22
Umidade do solo a 10kPa (15cm)	-14,72	16,02	-0,92	0,39	2,11

A Tabela 4 apresenta os resultados dos coeficientes estimados pela análise de regressão *stepwise* no sentido *backward*. Observa-se que a regressão *stepwise* eliminou do modelo as variáveis explicativas densidade das partículas do solo na camada superficial, condutividade hidráulica do solo a 15 e 60cm de profundidade e condutividade hidráulica do lençol freático, restando 9 variáveis explicativas: teor de argila a 60cm de profundidade; distância do poço até o exutório da bacia; densidade das partículas do solo a 60cm de profundidade; densidade média do solo na camada superficial do solo; densidade média do solo a 60cm de profundidade; teor de matéria orgânica na camada superficial do solo; porosidade do solo a 60cm de profundidade; teor de silte a 60cm de profundidade; e umidade do solo submetida à tensão de 10kPa a 60cm de profundidade.

O índice de inflação da variância (VIF) reduziu para todas as variáveis explicativas selecionadas pela análise de regressão *stepwise*, quando comparado aos resultados

obtidos para o *full model*. Nenhuma variável apresentou índice de inflação da variância maior do que 15 o qual foi o limite definido como aceitável por este estudo. Todas as variáveis explicativas apresentaram VIF menor do que 10, conforme sugerido por Souza (1998) como referência para indicação de multicolinearidade entre as variáveis, com exceção da densidade média do solo a 60cm de profundidade que apresentou VIF igual a 13,35.

Verifica-se nos resultados expostos na Tabela 4 que, diferentemente dos resultados obtidos para o *full model*, seis das nove variáveis selecionadas tiveram p-valor menor do 0,05 na estimativa dos coeficientes pela regressão: teor de argila a 60cm de profundidade; densidade das partículas do solo a 60cm de profundidade; densidade média do solo na camada superficial do solo; densidade média do solo a 60cm de profundidade; porosidade do solo a 60cm de profundidade; e teor de silte a 60cm de profundidade. Como três variáveis explicativas não apresentaram significância estatística (distância do poço até o exutório da bacia; teor de matéria orgânica na camada superficial do solo; e umidade do solo submetida à tensão de 10kPa a 60cm de profundidade), optou-se por testar a regressão linear com a eliminação dessas três variáveis.

Tabela 4: Resultado dos coeficientes estimados na regressão para as variáveis selecionadas para compor o modelo selecionado pela regressão *stepwise* no sentido backward.

Variáveis explicativas	Estimativa do coeficiente	Erro Padrão	t-valor	p-valor	VIF
Constante	-78,47	22,50	-3,49	0,01	-----
Teor de argila (60cm)	0,12	0,04	2,74	0,02	5,16
Distância do poço até o exutório	-0,0002	0,00	-1,42	0,18	5,15
Densidade das partículas do solo (60cm)	4,91	2,04	2,41	0,03	2,01
Densidade média do solo (camada superficial do solo)	6,82	2,16	3,15	0,01	1,58
Densidade média do solo (60cm)	25,34	8,67	2,92	0,01	13,35
Teor de matéria orgânica (camada superficial do solo)	-0,34	0,27	-1,24	0,24	1,53
Porosidade do solo (60cm)	65,35	16,30	4,01	0,00	5,98
Teor de silte (60cm)	0,20	0,07	2,95	0,01	8,17
Umidade do solo a 10kPa (15cm)	-14,70	12,68	-1,16	0,27	1,98

O resultado do teste F obtido pela ANOVA realizada para comparar estes dois modelos (com nove e com seis variáveis explicativas) teve p-valor igual a 0,35, indicando que a hipótese nula em que os dois modelos são iguais não pode ser rejeitada. Dessa forma, como se busca um modelo linear com maior parcimônia, ou seja, com o mínimo de variáveis explicativas, optou-se por avaliar todos os resultados do modelo com seis variáveis explicativas para verificar se esse modelo pode ser usado para explicar a relação entre os fatores ambientais e a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim no período entre 17/02/2004 e 10/01/2012.

Os resultados da análise de regressão do modelo com seis variáveis explicativas para explicar a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim teve  $R^2$  igual 0,84,  $R^2$  ajustado para 14 graus de liberdade igual a 0,73, e AIC igual a 59,52. O resultado do *Akaike Information Criterion* “corrigido”, AICc, para esse modelo foi igual a 71,52. O teste F para esse modelo apresentou p-valor igual a 0,000192, indicando que a hipótese nula de que todas as variáveis explicativas são iguais pode ser rejeitada ao nível de significância de 1%.

A Tabela 5 apresenta os resultados dos coeficientes estimados pela análise de regressão para o modelo com maior parcimônia. Novamente o índice de inflação da variância (VIF) reduziu em relação aos resultados dos outros modelos apresentados nas Tabela 3 e Tabela 4. Desse modo, nenhuma variável apresentou índice de inflação da variância maior do que 15 o qual foi o limite definido como aceitável por este estudo. Todavia, a variável explicativa densidade média do solo a 60cm de profundidade apresentou VIF igual a 12,29, permanecendo ligeiramente maior do que o sugerido por Souza (1998) como referência para indicação de multicolinearidade entre as variáveis.

Verifica-se nos resultados expostos na Tabela 5 que, diferentemente dos resultados obtidos para o *full model* e para o modelo selecionado pela regressão *stepwise*, todas as seis variáveis selecionadas tiveram p-valor menor do 0,05 na estimativa dos coeficientes pela regressão. Após a avaliação dos resultados até aqui demonstrados, optou-se por realizar a análise dos resíduos do modelo com maior parcimônia, cujos resultados estão na Tabela 5, para assim validar o uso deste modelo para explicar a relação entre a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim no período entre 17/02/2004 e 10/01/2012 e os fatores ambientais existentes na respectiva bacia.



Tabela 5: Resultado dos coeficientes estimados na regressão para as variáveis selecionadas para compor o modelo com maior parcimônia (seis variáveis explicativas).

Variáveis explicativas	Estimativa do coeficiente	Erro Padrão	t-valor	p-valor	VIF
Constante	-94,34	20,70	-4,56	0,0004	-----
Teor de argila (60cm)	0,11	0,04	2,64	0,0194	5,13
Densidade das partículas do solo (60cm)	6,37	1,74	3,67	0,0025	1,39
Densidade média do solo (camada superficial do solo)	7,46	2,18	3,42	0,0042	1,54
Densidade média do solo (60cm)	25,12	8,51	2,95	0,0105	12,29
Porosidade do solo (60cm)	72,31	15,58	4,64	0,0004	5,23
Teor de silte (60cm)	0,21	0,06	3,40	0,0043	6,74

O teste Koenker (*Koenker's studentized Breusch-Pagan statistic*) para avaliar a estacionariedade do modelo teve p-valor igual 0,57 com 6 graus de liberdade, e índice igual a 4,76, indicando que a hipótese nula,  $H_0$ , que representa a situação em que o modelo é estacionário não pode ser rejeitada. Assim, a relação entre as variáveis explicativas e a variável resposta se mantém constante com a mudança na magnitude das variáveis explicativas, indicando homocedasticidade dos resíduos. A Figura 2 apresenta a relação entre os resíduos e os valores ajustados pela regressão reforçando o resultado obtido pelo teste Koenker.

O teste de Jarque-Bera para avaliar a normalidade dos resíduos apresentou p-valor igual a 0,64 e índice igual 0,90. O teste de Shapiro-Wilks, por sua vez, apresentou resultado igual a 0,97 e p-valor igual a 0,73. Para ambos os testes a hipótese nula que assume que a distribuição de probabilidade da variável aleatória de interesse se assemelha à distribuição normal não pode ser rejeitada. A Figura 3 apresenta a relação entre os quantis teóricos da distribuição normal e os quantis dos resíduos da regressão, QQnorm, reforçando o resultado obtido pelo teste de Jarque-Bera pelo teste de Shapiro-Wilks.

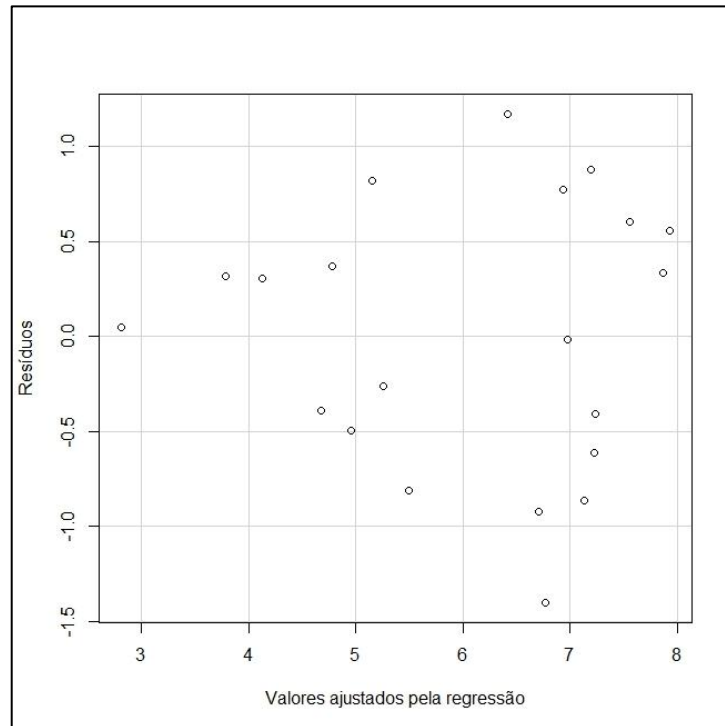


Figura 2: Relação entre os resíduos e os valores ajustados pela regressão do modelo com maior parcimônia (seis variáveis explicativas).

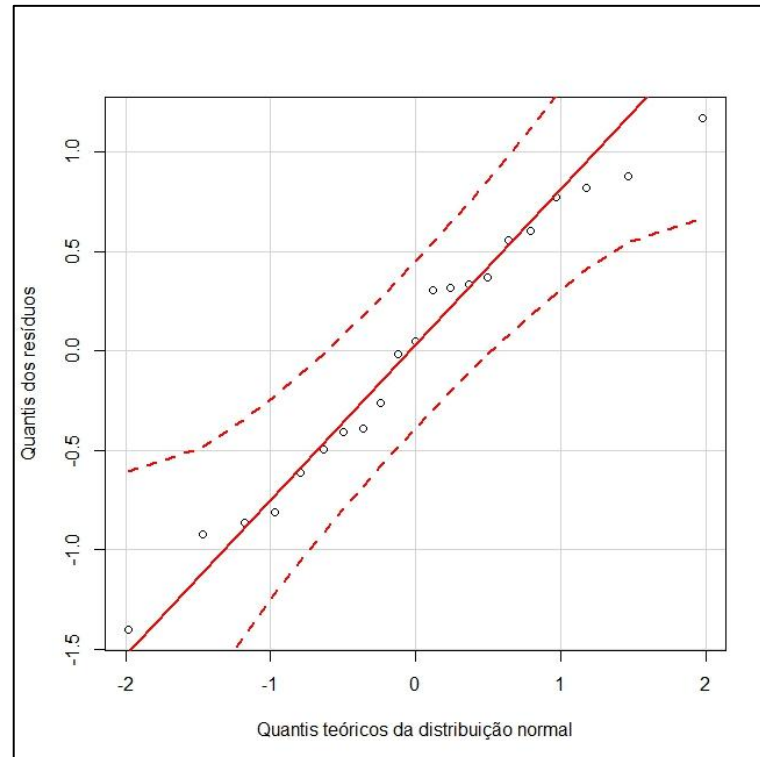


Figura 3: Gráfico QQnorm dos resíduos do modelo com maior parcimônia (seis variáveis explicativas).

Em relação à existência de valores extremos que podem influenciar tanto positivamente quanto negativamente nos resultados da análise de regressão, o teste de Bonferonni não identificou nenhum resíduo com p-valor de Bonferonni menor do que 0,05 de modo que a hipótese nula que indica que não há *outliers* nos resíduos não pode ser rejeitada. O poço piezométrico nº 31 apresentou o resíduo mais extremo, -1,40m (ou -1,94m em relação ao *t-student*), todavia o p-valor para o teste de Bonferonni para esse poço foi igual 0,07, o que não permite identificá-lo como *outlier*, conforme exposto anteriormente.

Quanto à autocorrelação espacial dos resíduos do modelo selecionado, o Índice de Moran calculado foi igual a -0,062 enquanto o Índice de Moran esperado foi igual a -0,05. A variância dos resultados nesse teste foi igual a 0,09 e o *z-score* foi igual a -0,04. Por fim, o p-valor para o Índice de Moran foi igual a 0,97 indicando que a hipótese nula que a variável de interesse se distribui aleatoriamente no espaço não pode ser rejeitada, e assim, não há autocorrelação espacial nos resíduos ao nível de significância de 5%.

Os resultados até aqui apresentados indicam que o modelo linear desenvolvido com as seis variáveis explicativas (teor de argila a 60cm de profundidade; densidade das partículas do solo a 60cm de profundidade; densidade média do solo na camada superficial do solo; densidade média do solo a 60cm de profundidade; porosidade do solo a 60cm de profundidade; e teor de silte a 60cm de profundidade) para correlacionar a amplitude máxima da variação do lençol freático na bacia do Alto Rio com os fatores ambientais existentes dentro da respectiva bacia hidrográfica, atendeu aos pressupostos necessários para evitar o que Souza (1998) definiu como “a utilização errônea das técnicas estatísticas clássicas”.

O VIF igual a 12,29 da variável explicativa densidade média do solo a 60cm de profundidade no modelo criado com as seis variáveis explicativas foi a exceção dos itens avaliados para validar o modelo desenvolvido, não obstante, conforme exposto anteriormente, optou-se por aceitar esse resultado. Além disso, as seis variáveis selecionadas explicaram a variação da amplitude máxima da variação do lençol freático em 73% o que representa um bom resultado, tendo em vista que não foi utilizada nenhuma variável hidrológica que compõem o balanço hidrológico, como por exemplo, a precipitação, a evapotranspiração e o escoamento superficial.

Entretanto, o modelo de regressão linear multivariado é considerado como um modelo com estimativas globais não captando as variações locais de cada variável (Fotheringham *et al.*, 2002). Desse modo, foi conduzida a análise de regressão com pesos espaciais a fim de incluir na modelagem a variabilidade espacial das variáveis utilizadas para correlacionar a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim com os fatores ambientais existentes dentro da respectiva bacia hidrográfica.

#### 4.3 Regressão com pesos espaciais

A Tabela 6 apresenta o valor mínimo, o valor do primeiro quartil, a mediana, o terceiro quartil, e o valor máximo dos coeficientes estimados na regressão com pesos espaciais. Observa-se que houve grande variabilidade nos coeficientes estimados na regressão com pesos espaciais. As seis variáveis explicativas e a constante tiveram tanto relações negativas quanto relações positivas com a variável resposta nos diferentes pontos amostrados.

Tabela 6: Sumário da estimativa dos coeficientes obtidos na regressão com pesos espaciais.

Variável explicativa	Mínimo	Primeiro quartil	Mediana	Terceiro quartil	Máximo
Constante	-277,64	-159,53	-48,68	21,32	100,98
Teor de argila (60cm)	-0,32	-0,16	0,07	0,41	0,87
Densidade das partículas do solo (60cm)	-0,94	1,16	7,94	8,67	9,25
Densidade média do solo (camada superficial do solo)	-12,97	5,58	6,77	17,79	47,06
Densidade média do solo (60cm)	-62,33	-4,32	2,62	31,29	44,50
Porosidade do solo (60cm)	-19,41	22,25	44,69	108,69	188,65
Teor de silte (60cm)	-0,39	-0,11	0,04	0,36	0,49

Com relação aos coeficientes das variáveis explicativas obtidos na regressão global, Tabela 5, as estimativas foram todas positivas, com exceção da constante que foi negativa. Isso demonstra que a regressão com pesos espaciais conseguiu captar relações que foram generalizadas como positivas na estimativa dos coeficientes, ao captar relações inversas entre a mesma variável explicativa e a variável resposta. Observa-se

também, que a estimativa dos coeficientes das variáveis explicativas na regressão global está dentro da amplitude interquartílica da estimativa dos coeficientes na regressão com pesos espaciais.

O diagnóstico da regressão com pesos espaciais do modelo resumido na Tabela 6 teve  $R^2$  igual 0,99,  $R^2$ ajustado igual a 0,90, e AICc igual a -1.220,71. Segundo Fotheringham *et al.* (2002), se a regressão global apresenta AICc menor do que o obtido na regressão com pesos espaciais, o componente espacial inserido no modelo não trouxe nenhuma vantagem ao modelo desenvolvido com os pesos espaciais. Conforme exposto anteriormente, o AICc calculado para o modelo de regressão global com seis variáveis explicativas apresentadas na Tabela 5 foi igual a 71,52. O teste F para esse modelo apresentou p-valor igual a 0,009, indicando que a hipótese nula de que todas as variáveis explicativas são iguais pode ser rejeitada ao nível de significância de 1%.

Quanto à autocorrelação espacial dos resíduos do modelo gerado pela regressão com pesos espaciais, o Índice de Moran calculado foi igual a -0,10 enquanto o Índice de Moran esperado foi igual a -0,05. A variância dos resultados nesse teste foi igual a 0,07 e o *z-score* foi igual a -0,20. Por fim, o p-valor para o Índice de Moran foi igual a 0,84 indicando que a hipótese nula que a variável de interesse se distribui aleatoriamente no espaço não pode ser rejeitada, e assim, não houve autocorrelação espacial nos resíduos do modelo ao nível de significância de 5%.

A Figura 4 apresenta a distribuição espacial, após interpolação dos resultados pelo método IDW, dos coeficientes estimados pela regressão com pesos espaciais.

A constante do modelo, o teor de argila a 60cm de profundidade, a densidade média do solo a 60cm de profundidade e o teor de silte a 60cm de profundidade apresentaram amplitude interquartílica, respectivamente, entre: -159,53 e 21,32; -0,16 e 0,41; -4,32 e 31,29; e 0,11 e 0,36.

Já a amplitude interquartílica dos coeficientes estimados para a densidade das partículas do solo a 60cm de profundidade, a densidade média do solo na camada superficial e a porosidade do solo a 60cm de profundidade variou, respectivamente, entre: 1,16 e 8,67; 5,58 e 17,79; e 22,25 e 108,69. A estimativa dos valores mínimos dos coeficientes para estas três últimas variáveis explicativas foram todas negativas sendo, respectivamente iguais a -0,94, -12,97 e -19,41.

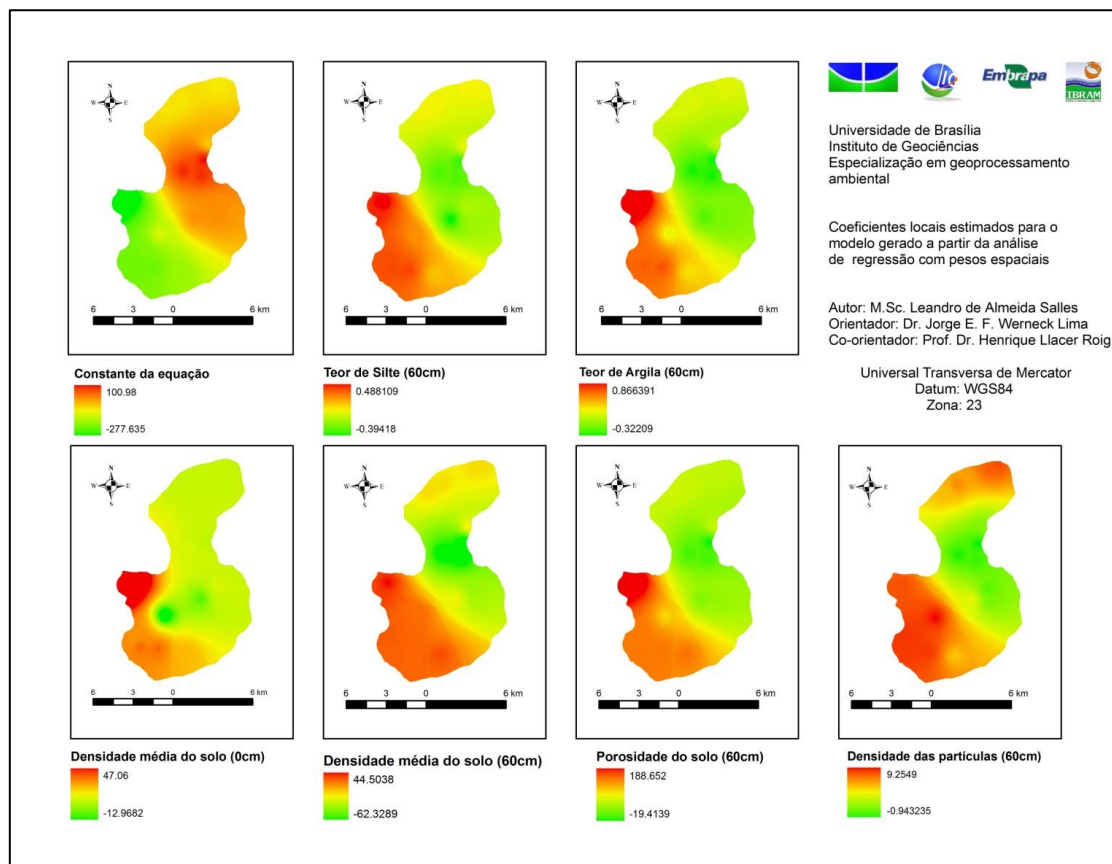


Figura 4: Distribuição espacial dos coeficientes estimados pela regressão com pesos espaciais.

Observa-se que, com exceção da constante do modelo, houve uma tendência de ocorrer maiores correlações na porção sul/sudoeste da bacia. Quanto às correlações negativas, observa-se que elas ocorreram predominantemente na região central da bacia, principalmente, na parte mais fina, novamente com exceção da constante do modelo. A densidade média do solo na camada superficial concentrou as correlações negativas abaixo da região central, saindo sutilmente do padrão das demais variáveis explicativas.

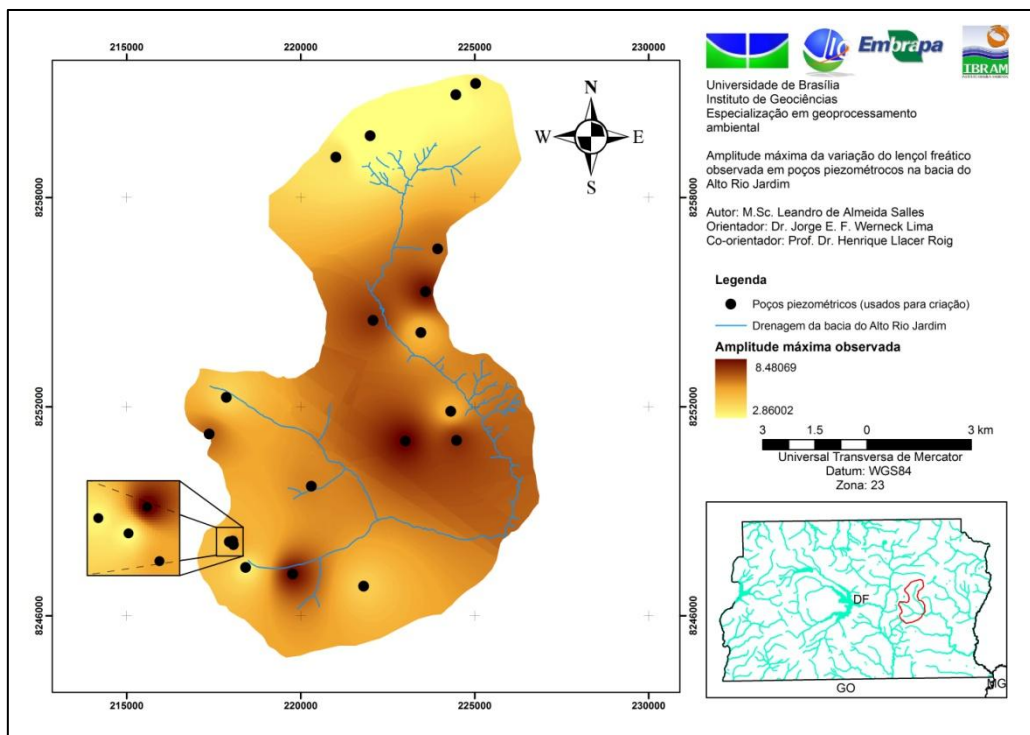
As Figura 5 (a), (b) e (c), por sua vez, apresentam, respectivamente, o valor observado da amplitude máxima de variação do lençol freático na bacia do Alto Rio Jardim no período entre 17/02/2004 e 10/01/2012, o valor estimado da amplitude máxima da variação do lençol freático, e os resíduos obtidos com modelo gerado pela regressão com pesos espaciais, após interpolação dos resultados pelo método IDW.

Observa-se que as estimativas da amplitude máxima da variação do lençol freático pelo modelo linear desenvolvido com pesos espécies foram bem semelhantes aos valores observados nos poços piezométricos.

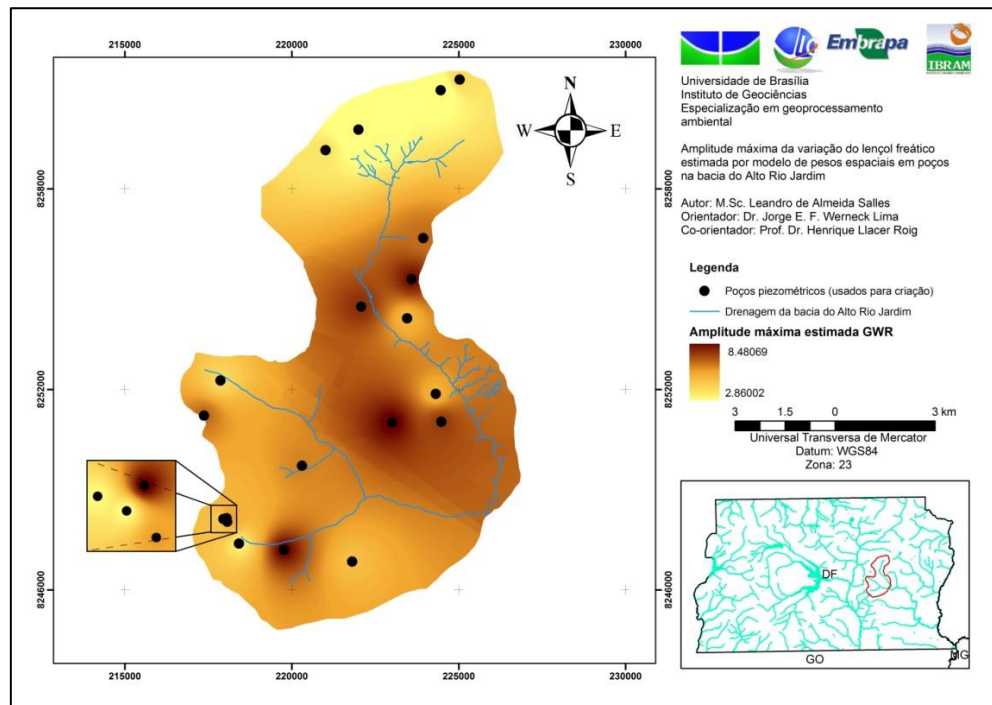
Os resíduos gerados nas estimativas do modelo tendem a se concentrar próximo ao divisor de águas da bacia do Alto Rio Jardim.

Na porção norte da bacia houve uma região em que os resíduos variaram bruscamente não ocorrendo uma transição entre o que o modelo superestimou e o que foi subestimado. Na parte sudeste da bacia, essa transição foi mais sutil, sendo dividida pela rede de drenagem da bacia.

(a)



(b)



(c)

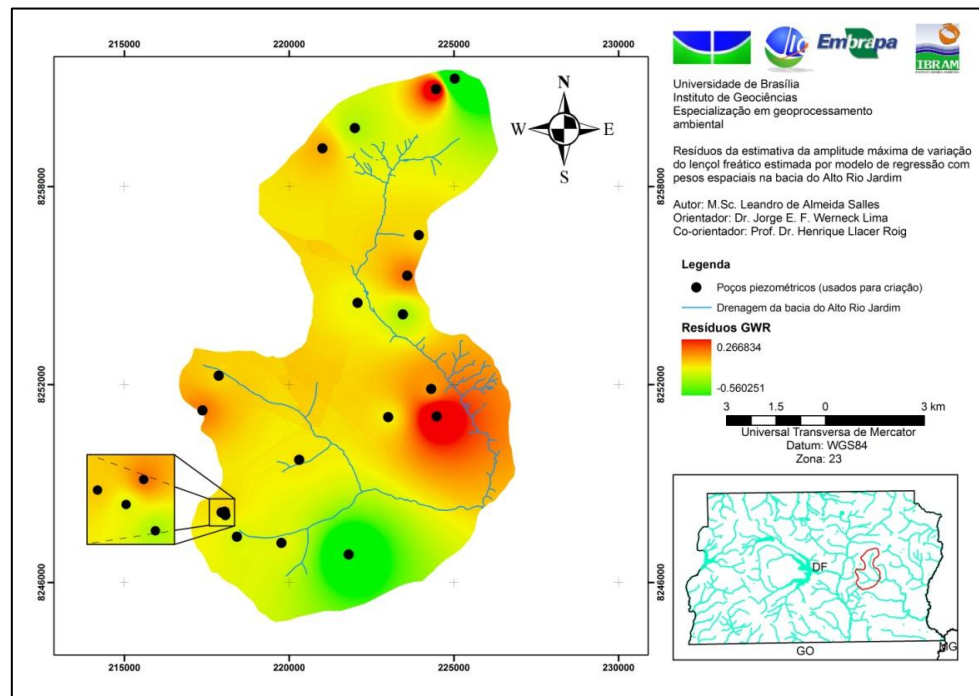


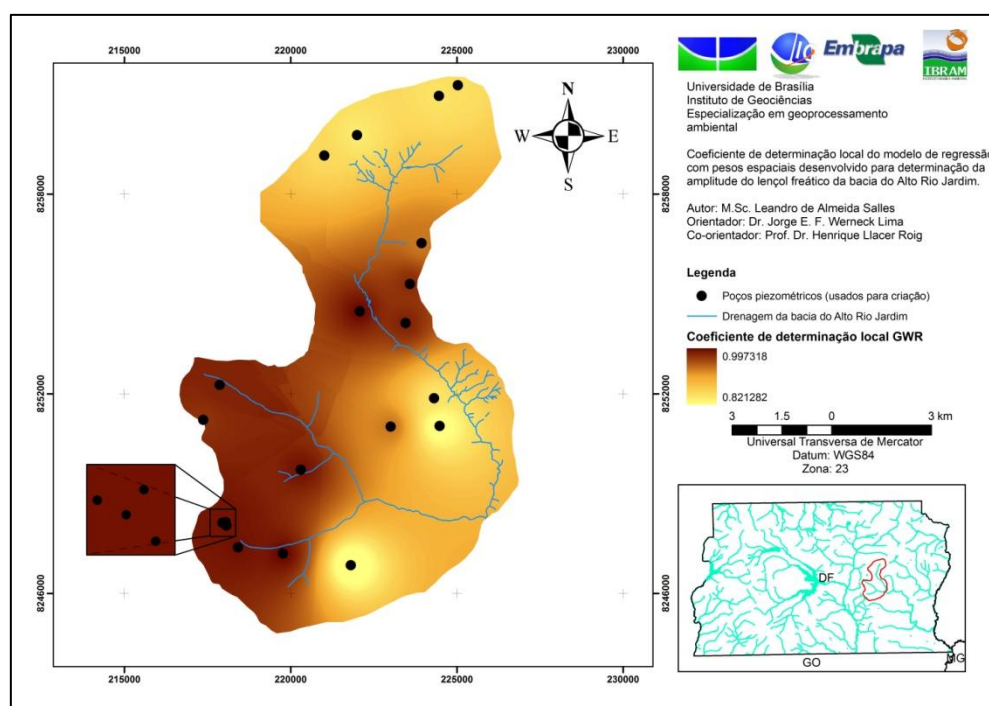
Figura 5: Amplitude máxima de variação do lençol freático na bacia do Alto Rio Jardim observada e estimada pela regressão com pesos espaciais, e respectivo resíduo do modelo (a), (b) e (c).



As Figura 6 (a) e (b) apresentam o  $R^2$  e o  $R^2$  ajustado do modelo gerado pela regressão com pesos espaciais para estimar a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim, após interpolação dos resultados pelo método IDW.

Observa-se que localmente, as melhores estimativas foram obtidas na porção sudoeste da bacia ocorrendo maiores valores do  $R^2$  e do  $R^2$  ajustado, enquanto na porção sudeste e setentrional ocorreram as piores estimativas do modelo.

(a)



(b)

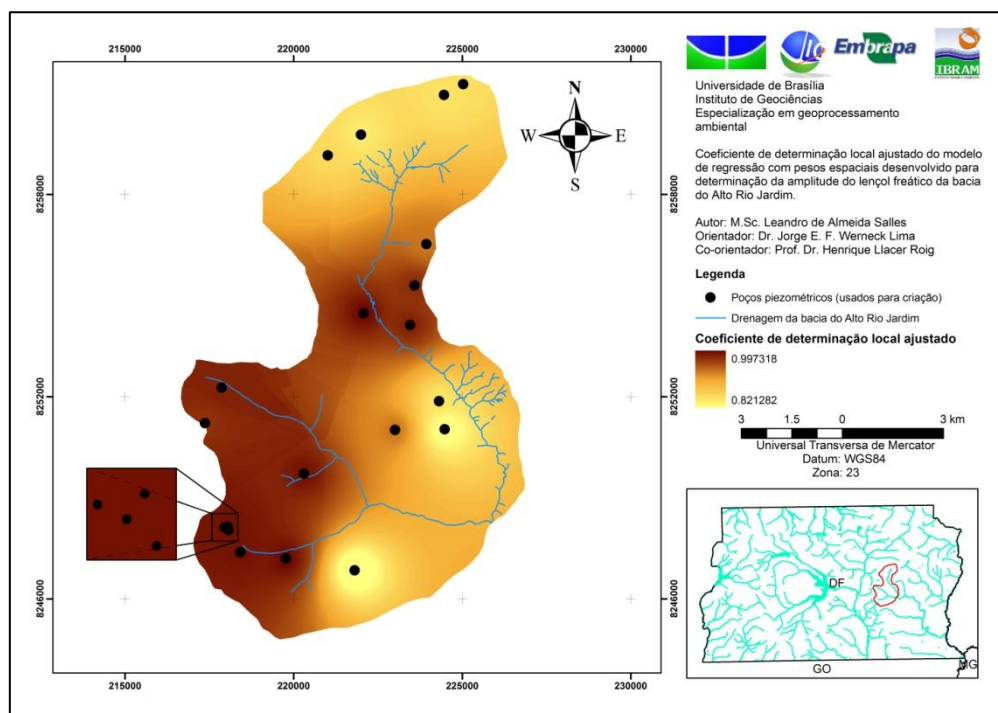


Figura 6:  $R^2$  local e  $R^2$  local ajustado do modelo gerado pela regressão com pesos espaciais para estimar a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim, (a) e (b).

Por fim, a Tabela 7 apresenta os resultados da análise de variância que comparou os resultados obtidos pelos modelos de regressão global e com pesos espaciais. Os resultados indicam que a hipótese nula não pode ser rejeitada, pois o p-valor foi igual a 0,56. Desse modo, não há diferença significativa entre as respostas dos modelos de regressão global e de pesos espaciais.

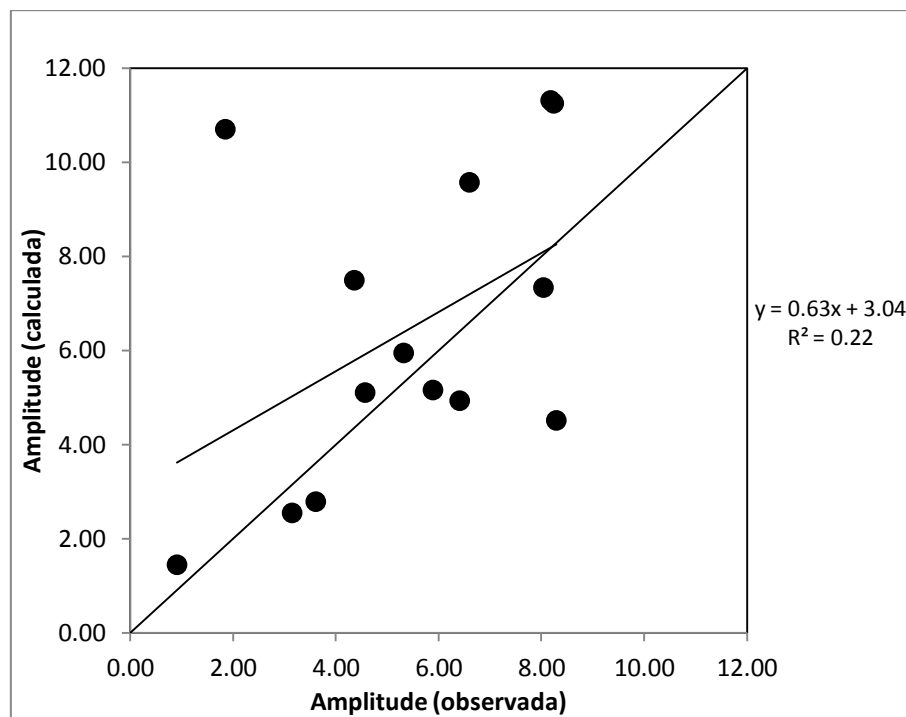
Tabela 7: ANOVA comparando os resultados do modelo de regressão global e com pesos espaciais.

Fonte de variação	SQ	GL	QM	F
Resíduos do modelo de regressão global	9,77	7,00		
Ganhos do modelo de regressão com pesos espaciais	9,17	12,72	0,72077	
Resíduos do modelo de regressão com pesos espaciais	0,60	1,28	0,46801	1,54005

#### 4.4 Validação dos modelos de regressão global e com pesos espaciais

As Figura 7 (a) e (b) apresentam os resultados da validação dos modelos de regressão global e com pesos espaciais desenvolvidos para simular a amplitude máxima da variação do lençol freático observada nos 14 poços piezométricos localizados na bacia do Alto Rio Jardim com os 14 poços piezométricos que foram separados para essa análise. A correlação entre os valores estimados e observados pelo modelo de regressão global e com pesos espaciais tiveram o  $R^2$ , respectivamente, iguais a 0,22, e 0,04.

(a)



(b)

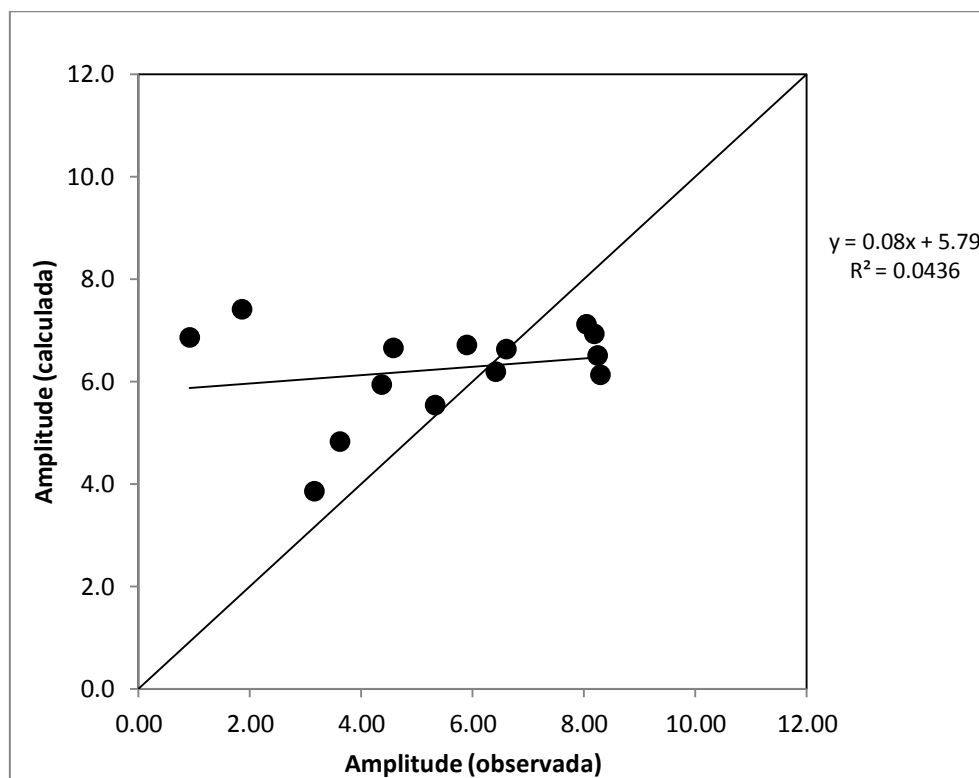
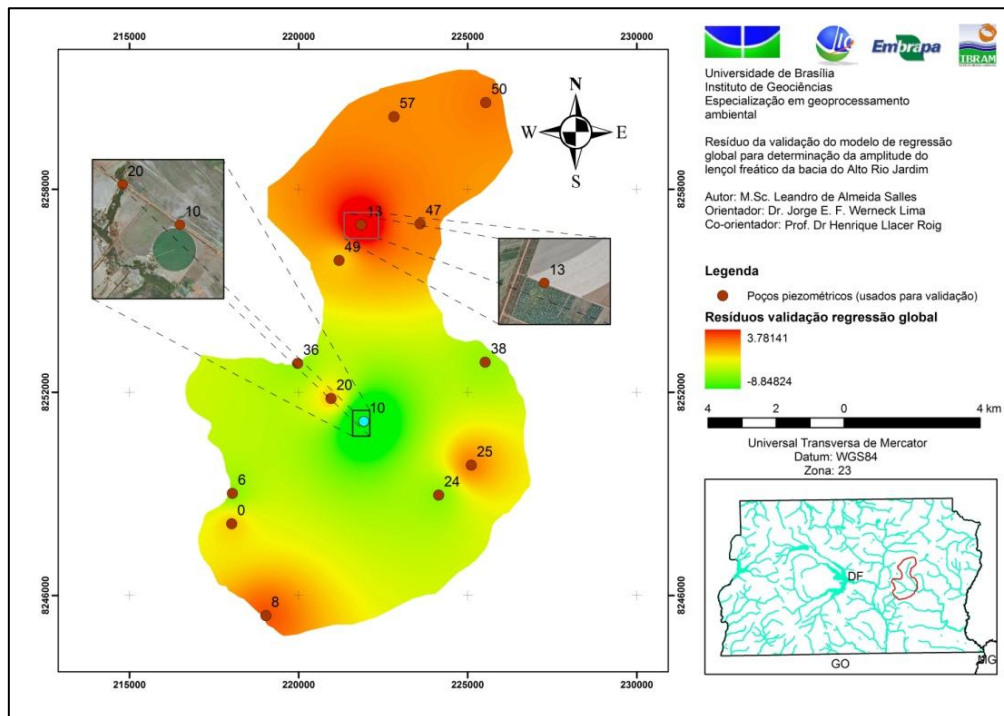


Figura 7: Correlação entre a amplitude máxima da variação do lençol freático observada nos poços piezométricos selecionados para validação localizados na bacia do Alto Rio Jardim e a estimada pelo modelo de regressão linear global (a) e com pesos espaciais (b).

A Figura 8 (a) e (b), por sua vez, apresentam os resíduos obtidos na validação dos modelos de regressão global e com pesos espaciais desenvolvidos para simular a amplitude máxima da variação do lençol freático observada nos 14 poços piezométricos localizados na bacia do Alto Rio Jardim. Observa-se que a amplitude simulada foi superestimada intensamente no poço nº10 apresentando resíduo igual a -8,85 e -5,56 para o modelo de regressão global e com pesos espaciais, respectivamente.

Os modelos de regressão global e com pesos espaciais também tiveram dificuldade para simular amplitude do nível freático no poço nº 13 apresentando resíduo igual a -3,78 e -5,95 para o modelo de regressão global e com pesos espaciais, respectivamente.

(a)



(b)

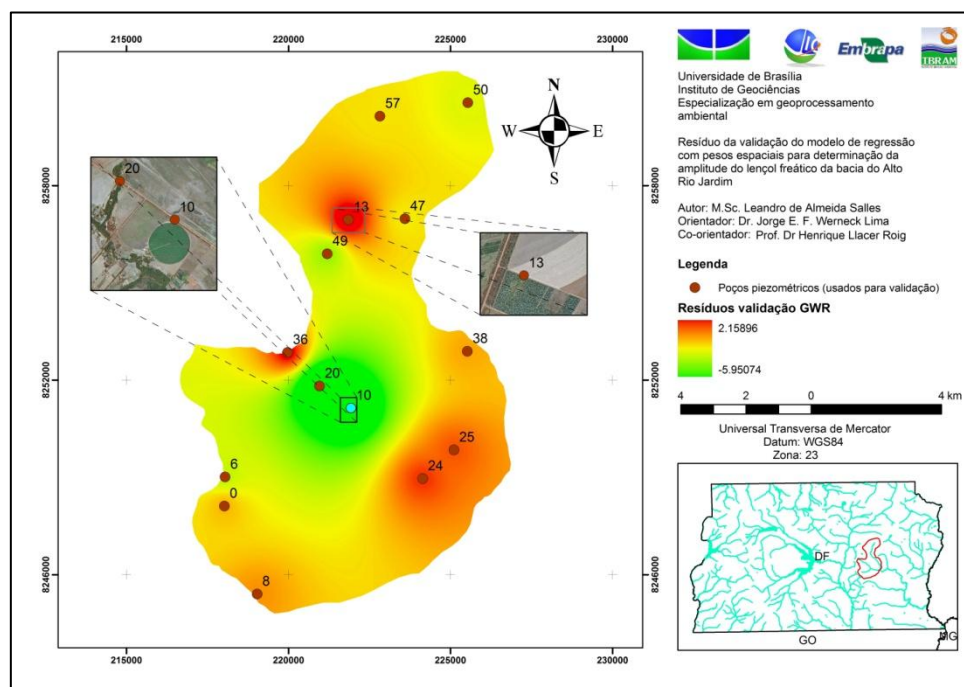


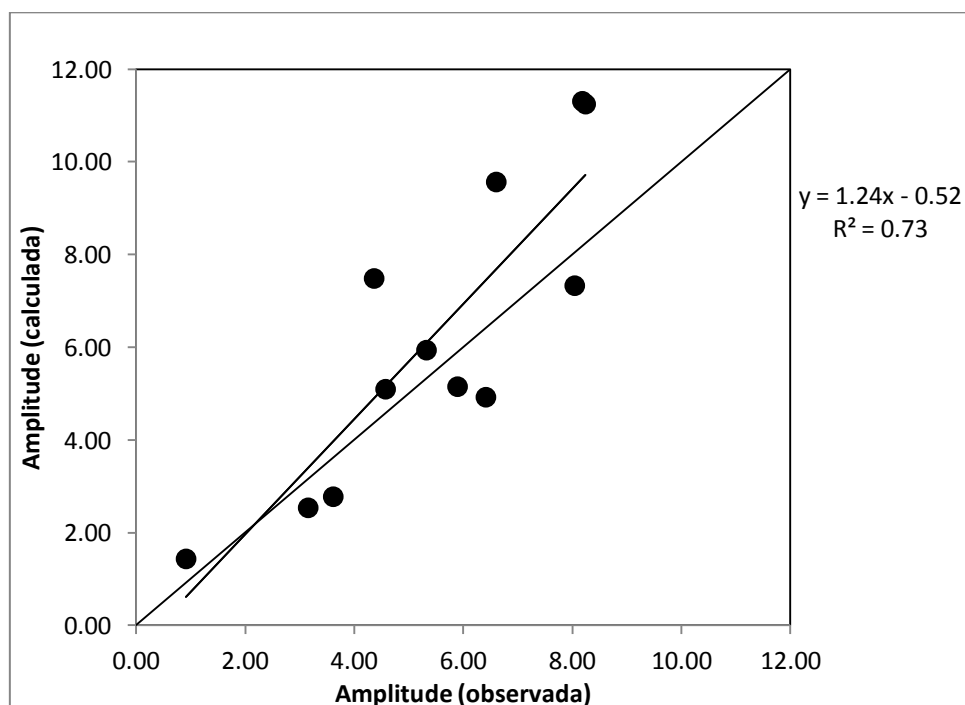
Figura 8: Resíduos obtidos na validação dos modelos desenvolvidos para simular a amplitude máxima da variação do lençol freático pela regressão global (a) e com pesos especiais e (b).

Observa-se também que modelo de regressão com pesos espaciais apresentou resíduo muito elevado na simulação do poço piezométrico nº 20, sendo igual a -5,95m, enquanto para o modelo de regressão global o resíduo neste local foi igual a -0,53m.

A Figura 8 também indica que havia um pivô central de aproximadamente 42ha próximo ao poço piezométrico nº10 em 2009, que pode ter interferido na flutuação do lençol freático. Em relação ao poço piezométrico nº13, é possível observar na Figura 8 que há um reflorestamento próximo a esse poço que também pode ter influenciado a amplitude máxima de variação do lençol freático por meio da evapotranspiração.

As Figura 9 (a) e (b) apresentam os resultados da validação dos modelos de regressão global e com pesos espaciais desenvolvidos para simular a amplitude máxima da variação do lençol freático observada retirando-se da análise os poços piezométricos nº10 e nº13. Observa-se que a correlação entre os valores estimados e observados pelo modelo de regressão global e com pesos espaciais aumentou significativamente, tendo  $R^2$ , respectivamente, iguais a 0,73, e 0,20.

(a)



(b)

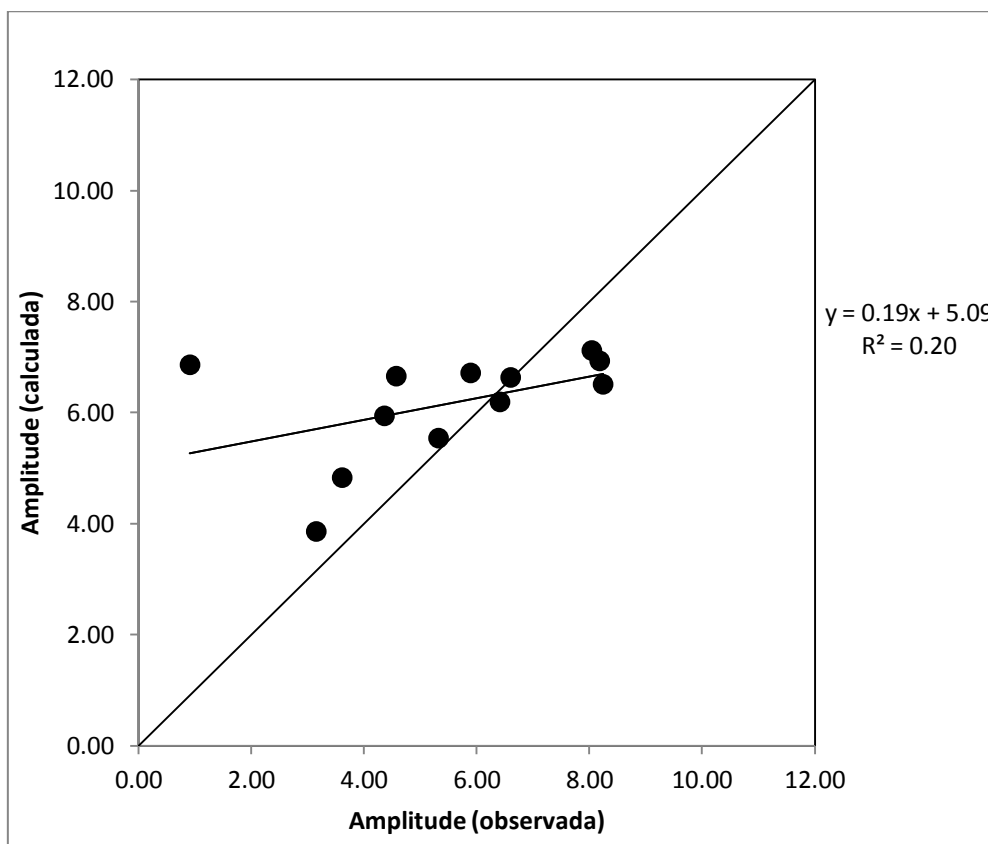


Figura 9: Correlação entre a amplitude máxima da variação do lençol freático observada nos poços piezométricos selecionados para validação, retirando-se os poços nº10 e nº13 localizados na bacia do Alto Rio Jardim e a estimada pelo modelo de regressão linear global (a) e com pesos espaciais (b).

A baixa correlação entre a amplitude máxima da variação do lençol freático estimada e observada pelo modelo de regressão com pesos espaciais após a retirada dos poços piezométricos nº10 e nº13 pode ser explicada devido ao resíduo calculado para a simulação da amplitude máxima de variação do lençol freático no poço piezométrico nº20 que, conforme exposto anteriormente, apresentou o maior valor residual (-5,95m) com o conjunto de dados selecionados para validação.

Retirando-se o poço nº20 que obteve o maior resíduo para o conjunto de dados selecionados para a validação dos resultados, além dos poços piezométricos nº10 e nº13, apenas para a validação do modelo gerado pela regressão com pesos espaciais, a correlação entre a amplitude máxima da variação do lençol freático estimada e

observada aumenta apresentando  $R^2$  igual a 0,63. Na Figura 10 apresenta-se esse resultado.

Observa-se na Figura 8(b) que o poço piezométrico nº 20 localiza-se próximo a uma mata de galeria. Nesse poço o modelo de regressão global conseguiu uma boa estimativa da amplitude máxima da variação do lençol freático com resíduo calculado muito pequeno (-0,53m), enquanto para o modelo com pesos espaciais foi a pior estimativa, com resíduo altíssimo (-5,95m).

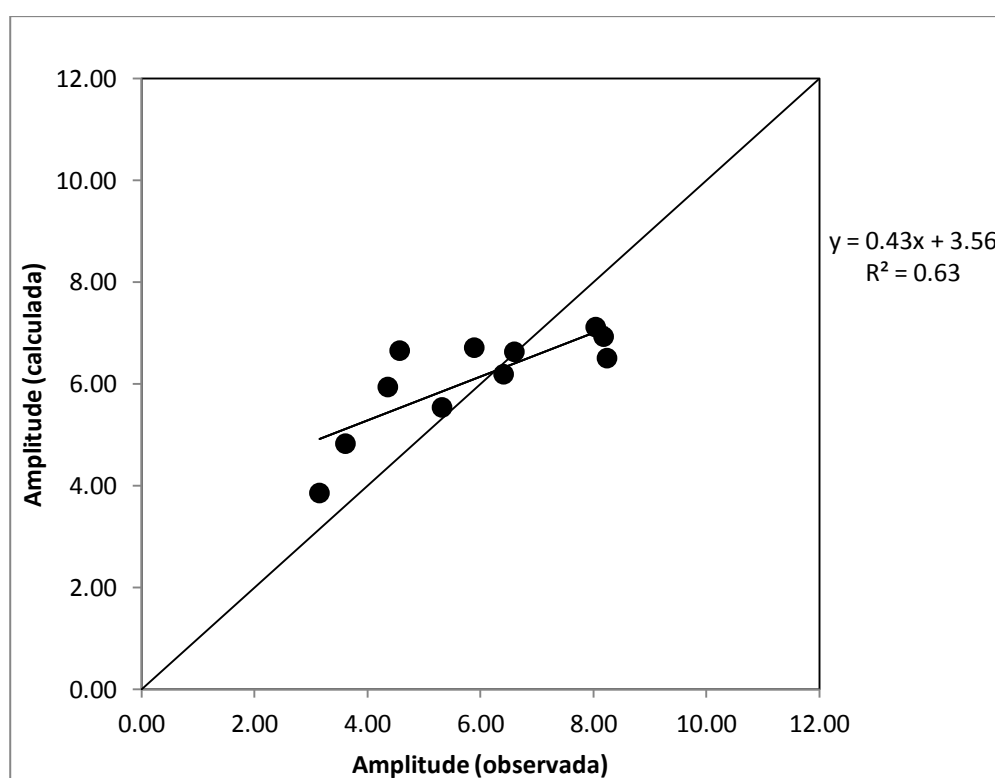


Figura 10: Correlação entre a amplitude máxima da variação do lençol freático observada nos poços piezométricos selecionados para validação, retirando-se os poços nº10, nº13 e nº20 localizados na bacia do Alto Rio Jardim e a estimada pelo modelo de regressão com pesos espaciais.

O modelo de regressão com pesos espaciais calcula para cada ponto uma equação de regressão linear utilizando o kernel espacial que é função do *bandwidth* utilizado, e que, por conseguinte, altera o número de graus de liberdade utilizados para os cálculos do



modelo local (Fotheringham *et al.*, 2002). Conforme exposto anteriormente o critério de seleção do *bandwidth* foi feito buscando aperfeiçoar o AICc e contemplando entre 10 e 45% das  $n$  observações vizinhas do  $i$ -ésimo ponto. A regressão global, por sua vez, foi estimada para 14 graus de liberdade. Assim, por esse motivo a regressão com pesos espaciais pode não ter conseguido filtrar bem as relações entre a variável resposta e as variáveis explicativas.

Lima (2010) levantou algumas hipóteses para explicar os problemas em que encontrou no processo de modelagem que podem ser estendidos para o presente estudo como: a representação inadequada do meio físico, o que inclui a precisão e a representatividade dos dados medidos; a representação inadequada do sistema de circulação da água na bacia, como a existência de fraturas ou caminhos preferenciais na zona do solo assumida como “impermeável”; e a não consideração da variabilidade vertical das características físico-hídricas dos solos dentro de cada camada de representação do meio.

## 5. CONCLUSÕES

A técnica espacial e não espacial de análise multivariada utilizadas no presente estudo foram capazes de estabelecer a correlação entre o comportamento do lençol freático da bacia hidrográfica do na bacia do Alto Rio Jardim no período entre 17/02/2004 e 10/01/2012 e as propriedades físicas dos solos existentes bacia.

Ambas as técnicas utilizadas apresentaram índices estatísticos que podem ser considerados satisfatórios na explicação da amplitude máxima da variação do lençol freático nos poços piezométricos localizados na bacia do Alto Rio Jardim pelo modelo.

A regressão com pesos espaciais conseguiu distribuir no espaço as correlações entre a amplitude máxima da variação do lençol freático na bacia do Alto Rio Jardim e as propriedades físicas dos solos selecionadas para explicar, todavia, não houve diferença significativa nos resultados das técnicas empregadas.

## 6. CONSIDERAÇÕES FINAIS

Os resultados apresentados no presente estudo buscaram atender as premissas básicas da estatística clássica. Todavia, o índice de inflação da variância que indica a existência de multicolinearidade entre as variáveis ficou ligeiramente acima do recomendado pela literatura. Assim, sugere-se que a pesquisa continue no sentido de buscar desenvolver um modelo cujas variáveis explicativas não possuam tanta correlação entre elas.

Além disso, vale ressaltar que não foi avaliado de forma efetiva na análise de regressão com pesos espaciais o *Bias-Variance Trade-Off* que pode ocorrer na seleção da *bandwidth* do kernel o qual é utilizado para estabelecer as relações de vizinhança entre as diferentes variáveis explicativas do modelo. Segundo Fotheringham *et al.* (2002), na medida em que a *bandwidth* do kernel aumenta, diminui o número de graus de liberdade utilizados para os cálculos do modelo o que pode aumentar a tendência (*Bias*) dos resultados do modelo. Também será interessante conduzir uma análise da estacionariedade individual de cada parâmetro, com técnicas de Monte Carlo.

## Referências Bibliográficas

BAYRAMOV, E.; BUCHROITHNER, M.F.; MCGURTY, E. (2012). **Determination of main climate and ground factors controlling vegetation cover regrowth along oil and gas pipelines using multiple, spatial and geographically weighted regression procedures.** Environmental Earth Sciences, 66: 2047-2062.

BIERMAN, P.; LEWIS, M.; OSTENDORF, B.; TANNER, J. (2011). **A review of methods for analyzing spatial and temporal patterns in coastal water quality.** Ecological Indicators, 11: 103-114.

BLAINEY, S. (2010). **Trip end models of local rail demand in England and Wales.** Journal of transport geography, 18: 153-165.

BREETZKE, G.D. (2012). **The effect of altitude and slope on the spatial patterning of burglary**. *Applied Geography*, 34: 66-75.

CHOW, V.T.; MAIDMENT, D.R.; MAYS, L.W. **Applied hydrology**. 1988. New York: McGraw-hill Book, 572p.

DELLER, S. (2011). **Spatial heterogeneity in the wage curve**. *Economics letters*, 113: 231-233.

DOOGE, J.C.I. (1973). **Linear theory of hydrologic systems**. Agricultural Research Service: Technical Bulletin n°1468, 327p.

ERDOGAN, S. (2010). **Modelling the spatial distribution of DEM error with geographically weighted regression: An experimental study**. *Computers & Geosciences*, 36: 34-43.

FOTHERINGHAM, A.S.; BRUNSDON, C.; CHARLTON, M. (2002). **Geographically Weighted Regression: the analysis of spatially varying relationships**. Chichester: John Wiley & Sons Ltd, 269 p.

FOX, J. (2005). **The R Commander: A Basic Statistics Graphical User Interface to R**. *Journal of Statistical Software*, 14(9): 1-42.

GAO, Y.; HUANG, J.; LI, S.; LI, S. (2012). **Spatial pattern of non-stationarity and scale-dependent relationships between NDVI and climatic factors – A case study in Qinghai-Tibet Plateau, China**. *Ecological Indicators*, 20:170-176.

GRUBESIC, T.H.; MACK, E.A.; KAYLEN, M.T. (2012). **Comparative Modeling approaches for understanding urban violence**. *Social Science Research*, 41:92-109.

GILBERT, A.; CHAKRABORTY, J. (2011). **Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida.** Social Science Research, 40: 273-286.

HARRIS, P.; BRUNDSON, C. (2010). **Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics.** Computers & Geosciences, 36: 54-70.

HIMMELBERGER, Y.O; PEARSALL, H.; RAKSHIT, R. (2009). **Concrete evidence & geographically weighted regression: A regional analysis of wealth and the land cover in Massachusetts.** Applied Geography, 29: 478-487.

IBEAS, A.; CORDERA, R.; DELL'OLIO, L.; MOURA, J.L. (2011). **Modelling demand in restricted parking zones.** Transportation research part A, 45: 485-498.

KUPFER, J.A; FARRIS, C.A. (2007). **Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models.** Landscape Ecology, 22: 837-852.

LANDIM, P.M.B. (2010). **Introdução à análise estatística de dados geológicos multivariados.** DGA, IGCE, UNESP/Rio Claro, Texto Didático 15, 229 pp., Disponível em <http://www.rc.unesp.br/igce/aplicada;textodi.html>. Acesso em 13/11/2012.

LANNA, A.E. (2001). **Elementos de estatística e probabilidade.** In: TUCCI, C.E.M. Hidrologia: ciência e aplicação. Segunda edição. Porto Alegre: Ed. Universidade/UFRGS: ABRH, p. 79-164.

LIMA, J.E.F.W.; EID, N.J.; SILVA, E.M. da.; MARTINS, E.S.; LOPES, W.T.A.; KOIDE, S. (2007). **Estimativa da taxa média anual de erosão na bacia experimental do Alto Rio Jardim – DF**. Anais do XVII Simpósio Brasileiro de Recursos Hídricos. São Paulo: ABRH. Cd-Rom.

LIMA, J.E.F.W. **Modelagem numérica do fluxo da água no solo e do escoamento de base em uma bacia experimental em área agrícola no Cerrado**. Tese de Doutorado, Publicação PTARH.TD – 08/10, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, DF, 312p. 2010

LEYK, S.; NORLUND, P.U.; NUCKOLS, J.R. (2012). **Robust assessment of spatial non-stationarity in model associations related to pediatric mortality due to diarrheal disease in Brazil**. Spatial and Spatio-temporal Epidemiology, 3: 95-105.

MITCHELL, A. (2005). **The ESRI guide to GIS analysis: volume 2 – Spatial Measurements & Statistics**. California: ESRI Press, 238 p.

MILLS, R.B.; PATERSON, A.M.; LEAN, D.R.S; SMOL, J.P.; MIERLE, G.; BLAIS, J.M. (2009). **Dissecting the spatial scales of mercury accumulation in Ontario lake sediment**. Environmental Pollution, 157: 2949-2956.

RANGEL, T.F.L.V.B; DINIZ-FILHO, J.A.F; BINI, L.M. (2010). **SAM: a comprehensive application for Spatial Analysis in Macroecology**. Ecography, 33:46-50. Disponível em <http://geodacenter.asu.edu/node/396#ans88>. Acesso em 23/11/2012.

REATTO, A.; CORREIA, J.R.; SPERA, S.T.; CHAGAS, C.S.; MARTINS, E.S.; AAANDAHUR, J.P.; GODOY, M.J.S.; ASSAD, M.L.C.L. (2000). **Levantamento semi-detalhado dos solos da bacia do rio Jardim-DF, escala 1:50.000**. Boletim de Pesquisa e Desenvolvimento 18, Planaltina: Embrapa Cerrados. 63p.

R DEVELOPMENT CORE TEAM (2012). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Disponível em <http://www.R-project.org>. Acesso em 28/09/2012.

SOUZA, G.S. (1998). **Introdução aos modelos de regressão linear e não-linear**. Brasília: Embrapa-SPI/Embrapa-SEA. 505p.

TUCCI, C.E.M. (1998). **Modelos e Simulações**. In: TUCCI, C.E.M. Modelos hidrológicos. Porto Alegre: Ed.Universidade/UFRGS/Associação Brasileira de Recursos Hídricos.

WU, WEI.; DRISCOLL, C.T. (2009). **Application of the PnET-BGC – An integrated biogeochemical model – To asses the surface water ANC recovery in the Adirondack region of New York under three multi-pollutant proposals**. Journal of hydrology, 378: 299-312